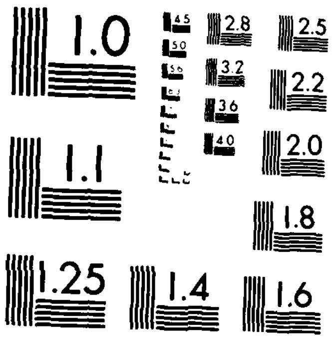
AD-A148 399 COMPATIBILITY EFFECTS AND PREFERENCE REVERSALS(U) 1/2
PERCEPTRONICS INC WOODLAND HILLS CA A TVERSKY ET AL.
21 AUG 84 PFTR-1127-84-8 N00014-82-C-0643
UNCLASSIFIED F/G 5/10 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

12

PFTR-1127-84-8
Contract No. N0014-82-C-0643
August 21, 1984

COMPATIBILITY EFFECTS AND PREFERENCE REVERSALS

AD-A148 399

Prepared for:

SCIENTIFIC OFFICER
OFFICE OF NAVAL RESEARCH
800 North Quincy Street
Arlington, VA 22217

PERCEPTRONICS

6271 VARIEL AVENUE • WOODLAND HILLS • CALIFORNIA 91367 • PHONE (213) 884-7470

PFTR-1127-84-8
Contract No. N0014-82-C-0643
August 21, 1984

COMPATIBILITY EFFECTS AND PREFERENCE REVERSALS

Prepared for:

SCIENTIFIC OFFICER
OFFICE OF NAVAL RESEARCH
800 North Quincy Street
Arlington, VA 22217

PERCEPTRONICS
6271 VARIEL AVENUE • WOODLAND HILLS • CALIFORNIA 91367 • PHONE (213) 884-7470

Abstract

This study investigates the effect of elicitation method on preferences among simple gambles. Three strategically equivalent elicitation procedures, choice, pricing, and attractiveness rating, produced reversals of preference when the same pairs of gambles were evaluated under different procedures. These results are attributed to the compatibility effect, a tendency to weight more heavily those aspects of the stimulus that are most easily mapped into the response. This phenomenon is described by a differential weighting model in which the effect of the elicitation procedure on the relative weighting of the stimulus attributes is expressed by a bias parameter b. Implications of these and related findings for the theory and the practice of decision making are discussed.

COMPATIBILITY EFFECTS AND PREFERENCE REVERSALS

Amos Tversky & Paul Slovic

Recent studies of decision making show that people's preferences among risky and riskless prospects often depend on the manner in which the options are described or framed (Kahneman & Tversky, 1979; Slovic, Fischhoff & Lichtenstein, 1982; Tversky & Kahneman, 1981). Much as changes in vantage point alter the apparent size of objects, different representations of a given decision problem induce predictable changes in preferences. These findings violate the normative principle of invariance, which states that the preference order between prospects should not depend on the manner in which they are described. That is, two versions of a choice problem that are recognized to be equivalent when shown together should elicit the same preference even when shown separately (Kahneman & Tversky, 1984; Tversky & Kahneman, 1984).

Invariance applies not only to the framing of options but to the elicitation of preferences as well: when preferences between options are expressed in several equivalent ways, each should produce the same ordering. However, invariance is often violated when preferences among gambles are elicited by different methods. The present study investigates the determinants of these failures of invariance, called preference reversals, and examines their implications for the theory and practice of decision making.

Background

The effect of elicitation method on preference between gambles was first observed by Slovic and Lichtenstein (1968), who found that both buying and selling prices for gambles were primarily determined by the dollar amounts that could be won or lost, whereas choices between gambles and ratings of their attractiveness were primarily influenced by the probabilities of winning and losing. Slovic and Lichtenstein reasoned that, if the method used to elicit preferences has differential effects on the weighting of the gamble's components, it should be possible to construct pairs of gambles such that the same individual would choose one member of the pair but set a higher price for the other. Lichtenstein and Slovic (1971, 1973) demonstrated such reversals in a series of studies, one of which was conducted on the floor of the Four Queens Casino in Las Vegas. A typical pair of gambles in the Las Vegas study consisted of a bet featuring a high probability of winning a modest sum of money (called the P Bet) and another bet featuring a low probability of winning a relatively large amount of money (called the $ Bet) as in the following example:

P Bet: 11/12 probability to win $3 and
1/12 probability to lose $6

$ Bet: 2/12 probability to win $19.75 and
10/12 probability to lose $1.25.

Each participant in this study first chose between the bets and later indicated a minimum selling price for each bet. For this pair of gambles, the two bets were chosen about equally often, but the

$ Bet received a higher selling price 88% of the time. Among the respondents who chose the P Bet, 87% gave a higher selling price to the $ Bet.

These findings have been replicated in numerous studies (see Hamm, 1984; Lindman, 1971; Mowen & Gentry, 1980; Pommerehne, Schneider, & Zweifel, 1982; Reilly, 1982; and a review by Slovic and Lichtenstein, 1983). A particularly careful replication was performed by Grether and Plott (1979), two skeptical economists who designed a series of experiments "to discredit the psychologist's works as applied to economics" (p. 623). Grether and Plott generated a list of 13 criticisms or potential artifacts that would render the preference reversal phenomenon irrelevant to economic theory. Their list included as possible explanations poor motivation, income effects, strategic responding, and the fact that the experimenters were psychologists (which might have led the respondents to be suspicious and behave peculiarly). Grether and Plott attempted to restore invariance by devising a special incentive system to heighten motivation and by controlling for possible biases. The study, of course, was conducted by economists. To their surprise, preference reversals remained much in evidence despite these determined efforts to eradicate them. It appears, then, that the discrepancy between choice and pricing is a highly robust phenomenon.

A recent study of preference reversals by Goldstein (1982) attempted to separate the effect of response mode (pricing vs. choosing) from the effect of stimulus presentation (single vs.

paired). To analyze the effect of these variables, Goldstein constructed four elicitation procedures. Two were single-stimulus methods: rating the attractiveness of a bet and setting its minimum selling price. Two were paired comparison methods: choosing between bets and ordering their minimum selling prices, without actually generating the prices. Goldstein found the usual reversals in which subjects chose the P bet over the $ bet but assigned a higher selling price to the $ bet. The comparison of choices with the ordering of selling prices did not produce many reversals. However, because subjects were required only to order the prices and not to state them, they may have simplified this task by treating it as a choice. Goldstein also observed that, unlike pricing , the rating response favored the P Bet over the $ Bet, yielding a new form of reversal: of the pairs in which the subjects gave a higher rating to the P Bet, the $ Bet received the higher selling price 65% of the time.

Hypotheses

Despite numerous experimental studies, the precise determinants of the effect are not entirely clear. The goal of the present study is to clarify the basis of preference reversals. We first investigate three specific hypotheses concerning the locus of this effect and then propose a more general explanatory mechanism.

The comparison hypothesis attributes preference reversals to the difference between pair-comparison and single-stimulus procedures. According to this account, P-bets are preferred to $-bets in a direct comparison, but the $-bets appear relatively more desirable when each bet is evaluated separately, using either a rating or a pricing procedure.

The generation hypothesis attributes reversals to the process of generating a selling price, or a cash equivalent, of a risky prospect. According to this account people overprice the bets because of anchoring and adjustment (Slovic & Lichtenstein, 1968; Tversky & Kahneman, 1974) or some other strategy. Consequently, people should prefer receiving the price they set over the opportunity of playing the bet.

The risk hypothesis attributes the reversal of preferences to the difference between a choice involving two bets and a choice involving a bet and a sure thing. According to this account the effect is due to the presence of a riskless option in pricing, not to the process of generating an explicit cash equivalent.

All three hypotheses are consistent with the basic discrepancies between choices and prices. Nevertheless, they lead to different predictions and they suggest different explanatory mechanisms. In particular, the comparison hypothesis locates the effect in the nature of the task (pair-comparison choice vs. single-stimulus evaluation); the generation hypothesis locates the effect in the nature of the response (choice vs. pricing); and the risk hypothesis locates the effect in the nature of the options (risky vs. riskless).

In order to test these hypotheses, the present study employed several variations of the preference reversal paradigm. First, following Goldstein (1982) and Slovic and Lichtenstein (1968), we included ratings of attractiveness along with choices and pricing as methods for eliciting preferences. The comparison of prices and ratings, both single-stimulus methods, provides a test of the comparison hypothesis. As a test of the generation hypothesis, we had subjects both generate prices for gambles and choose between these gambles and similar prices set by the experimenters. Third, to insure the strategic equivalence of our three elicitation procedures, we devised a method for linking preferences to outcomes that is identical across all conditions. Subjects were told that a pair of bets would be selected and the bet that received the higher attractiveness rating (or the higher price, or that was preferred in the choice task) would be the bet they would play. Consequently, there is no reason for the preferences elicited by prices and ratings to differ from each other or from the preferences elicited by direct choices.

Two additional aspects of the design were employed to minimize response-mode effects. One of these was to simplify the gambles by eliminating losses, so that each gamble consisted merely of a stated probability of winning a given amount and a complementary probability of winning nothing. The other was to actually play some of the gambles, to motivate careful evaluations.

Method

Subjects

The subjects were 189 people (72 men and 107 women) who responded to an advertisement in the University of Oregon student newspaper. They were paid $8 for participating in a 90-minute session that included several different experiments.

Stimuli

Six pairs of gambles, each containing one P bet and one $ bet, served as stimuli in this study (see Table 1). These gambles were obtained by deleting the losses from the six pairs studied by Lichtenstein and Slovic (1971; Experiment III), Grether and Plott (1979) and Goldstein (1982).

Insert Table 1 about here

Response Modes

Each subject was asked to evaluate the 12 gambles in Table 1 using two different response modes. Subjects in Group A (N = 94) rated the attractiveness of playing each of the 12 gambles. They also saw the gambles paired as in Table 1 and were asked to choose the gamble in each pair that they would prefer to play. Subjects in Group B (N = 63) evaluated each gamble individually in terms of its monetary worth and also made choices from each of the six pairs. Subjects in Group C (N = 32) evaluated each gamble both in terms of monetary worth and by rating its attractiveness.

In each group, about half of the subjects used one response mode first and immediately thereafter evaluated the bets with the second response mode. The remaining subjects used the two response modes in the reverse order.

Instructions

The instructions for each response condition began by introducing the particular concept of preference to be evaluated (attractiveness, choice, or monetary worth).

Attractiveness (Rating):

We're going to show you a number of bets. We would like you to rate how attractive each bet is to you. Imagine that two of these bets will be selected at random and that you will get to play the bet to which you gave the higher attractiveness rating, so your ratings of attractiveness will determine which bet you play.

For each of the bets on the following pages, make your rating of the bet's attractiveness by circling one number on the rating scale, which looks like this:

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

not at all an attractive bet	moderately attractive bet	an extremely attractive bet

Choice:

We're going to show you a number of pairs of bets. We would like you to indicate, for each pair, which bet you would prefer to play. Imagine that one of the pairs will be selected at random and that you will get to play the bet you preferred.

For each of the pairs on the following pages, indicate which bet you would choose to play. The answer sheet will look like this:

Bet A	Bet B
27/36 to win $2.50	6/36 to win $8.50

Mark one space:

A	____	____	____	____	B
	Strong Preference for A	Slight Preference for A	Slight Preference for B	Strong Preference for B	

As you make each choice, do it as though each one were the <u>only</u> choice you were going to make. Each choice should be made only on the merits of the two bets you are looking at—independently of any choices you have already made.

<u>Worth (Pricing)</u>:

We're going to show you a number of bets. For each bet we would like you to indicate how much the bet is worth to you. Later, two of these bets will be selected at random and you will get to play the bet you judged to be worth more to you, so your judgments of worth will determine which bet you play.

For each of the bets on the following pages, express your opinion about the bet's worth by stating an amount of money that is worth as much to you as is playing the bet.

What is the worth of the bet offering a 27/36 chance to win $2.50? Would you equate its worth with $2.45? That's probably too much. How about $.50? That's probably too little. Somewhere in between is the right amount such that you would find receiving that amount and playing the bet equal in worth. <u>Never put more than the bet's amount to win. That's the absolute maximum</u>.

A sample bet offering a chance of 27/36 to win $2.50 was displayed and the instructions explained how such a gamble would be played on a roulette wheel having 36 numbered sectors. The complementary 9/36 chance was described as leading to no win.

Subjects in all conditions were told that there were no right or wrong answers to the problems and that the investigators were only interested in their opinions about these bets. The remaining instructions were similar for all three response modes:

> After you finish both parts, we will randomly select 15% of the people in this room and give them the opportunity to actually play one of these bets. If you are selected to play, a pair of bets will be selected at random from one of the two parts and the bet that you
>
> rated as more attractive[1]
>
> chose to play[1]
>
> judged to be worth more[1]
>
> will be the bet you get to play.
>
> The bets will be played by spinning a roulette wheel. Those of you who win money will be able to keep your winnings. There are no losses. So make your judgments carefully. If you are selected, your preferences will determine which bet you play.

The bets were displayed in booklets with five or six bets, or pairs of bets, on each page. The order of presentation was fixed, with the two bets from pair 1 coming first, followed by the two bets from pair 2, etc. Within each pairing, the order of the P bet and

----

1. Depending on the response mode in effect for the selected bets.

the B bet was randomized. Two practice gambles were included in the attractiveness rating and the pricing conditions.

After setting prices and choosing among the six pairs, subjects in Group B were given each of the 12 gambles paired with a sure gain selected to be roughly as attractive as that gamble. The subjects were then asked to indicate, in each case, whether they preferred the gamble or the sure gain. This task was included to test whether the prices generated by the subjects were inflated due to insufficient adjustment from an anchor.

Each experimental group was run separately. Fifteen percent of the subjects were selected to play a bet. Those who won kept their winnings in addition to the payment for participating in the experiment.

Results

The order in which the response modes were employed had little effect on the results. Therefore the data reported here are combined across both orders within each group.

For each pair of gambles, subjects were classified according to whether they gave a higher attractiveness rating (or price) to the P bet or to the $ bet and whether they chose the P bet or the $ bet. Tied ratings and prices (about 7% of the comparisons) were excluded from the data analysis, with negligible effect on the results.

This classification produced the three sets of 2x2 matrices shown in Table 2. Inspection of these matrices reveals the influence of response modes on preferences. In choice, P bets were chosen over $ bets 64% of the time in Group A and 65% of the time in

Group B. This contrasts with pricing, in which P bets were given higher prices only 25% of the time in Group B and 13% of the time in Group C, and with ratings, in which P bets were rated more attractive 89% of the time in Group A and 86% of the time in Group B.

Insert Table 2 about here

The effects of response mode on the percentages of subjects preferring the P bet entail a substantial proportion of reversals within subjects. Across all pairs, the percentage of anticipated reversals was 27% for ratings vs. choice (Group A), 46% for pricing vs. choice (Group B), and 83% for rating vs. pricing (Group C). Reversals in the opposite direction occurred in only 2%, 6%, and 1% of the comparisons in Groups A, B, and C, respectively.

Table 3 presents the mean prices and ratings for each bet. Overall, the mean prices were more than 50% higher for the $ bets than for the P bets. In contrast, the mean ratings for the P bets were more than 50% higher than the means for the $ bets. Moreover, all P bets received higher mean attractiveness ratings than any of the $ bets. Further indication of the impact of response mode was the fact that the correlation between mean prices and mean bids was actually negative ($r = -.35$).

Insert Table 3 about here

Suppose subjects overpriced the bets, as implied by the generation hypothesis. When faced with a subsequent choice between receiving the inflated price or playing the gamble, the subject should choose the inflated price. Recall that the subjects in Group B set prices for all gambles and also chose between each of the gambles and a fixed sure-gain, denoted by X (see Table 4). If the bets are overpriced, because of insufficient adjustment or any other reason, the percentage of subjects who prefer the sure-thing X over the corresponding gamble (last column in Table 4) should be higher than the percentage of subjects whose stated prices were smaller than X (next-to-last column in Table 4). For example, suppose that the gamble and the sure gain X were each selected about 50% of the time. If the prices were inflated, then X should appear below the 50th percentile in the distribution of stated prices. Table 4 shows no systematic differences in these percentages for the $ bets and a slight difference in the opposite direction for the P bets, contrary to what might be expected if the bets were overpriced.

Insert Table 4 about here

Theoretical Analysis

In this study, we have investigated the effect of elicitation methods on preferences between simple risky prospects. We employed three strategically equivalent elicitation procedures, choice, pricing, and attractiveness rating, which led to markedly different preferences. The pricing response favored the $ bets while the rating response, and to a lesser extent the choices, favored the P bets. These differences produced many reversals of preference when the same pairs of gambles were compared under different procedures. In particular, we obtained the usual reversals between pricing and choice (observed by Lichtenstein & Slovic, 1971), as well as the reversals associated with attractiveness ratings that were observed by Goldstein (1982).

The three elicitation procedures employed in the present study are virtually identical because the ratings and the prices, like the choices, were used only to order the bets. The marked differences between the preferences induced by these procedures, therefore, cannot be explained by models (e.g., Fishburn, 1983; Loomes & Sugden, 1983) that attempt to rationalize preference reversals by extending the scope of the traditional normative theory. The finding that the ratings of attractiveness and the assessment of monetary worth yielded drastically different preference orders, with the choice being intermediate between the two, excludes the comparison hypothesis according to which preference reversals are due to the difference between single-stimulus evaluation (i.e., pricing and rating) and pair-comparison choice. The present results are also at variance with the generation hypothesis, which attributes reversals to overpricing. The results described in Table 4 show that the choices between bets and sure-things do not depart from the worth estimates, as implied by the generation hypothesis. Instead, the results appear to support a differential weighting model that is consistent with the risk hypothesis. It appears that the relative weight of payoffs to probabilities is larger in comparison of bets and cash amounts (whether specified by the subject or given by the experimenter) than in choices between bets, or in ratings of attractiveness. Across the 12 gambles, mean prices correlated .92 with payoffs and -.58 with probabilities; mean ratings correlated -.57 with payoffs and .95 with probabilities.

Why do payoffs loom larger in comparisons involving a cash amount and a gamble than in choices between bets or in ratings of attractiveness? We propose that differential weighting of the components of the gamble is controlled, in part at least, by the compatibility with the response. Compatibility can be viewed as the ease of coding or mapping the stimulus component into the response. The easier it is to execute such a mapping, we propose, the greater the weight given the component. Prices, or cash amounts, are clearly more compatible with payoffs than either ratings or choices are, because prices and payoffs are both expressed in dollars. The greater compatibility of ratings with probabilities may result from probabilities being more readily coded as attractive or unattractive than are payoffs. For example, 33 out of 36 chances to win are clearly attractive odds. On the other hand, a $4 payoff may be harder to code because the payoff component has no natural upper bound.

The compatibility effect has also been observed in other studies of judgment and choice. For example, Slovic and MacPhillamy (1974) asked subjects to predict, on the basis of test scores, which of two students, A or B, would get the higher grade point average in college. One test (available for both students), was common; the others were not. In the example below, the common test is English Skills. The other information was unique--Quantitative Ability for Student A and Need for Achievement for Student B.

	Student A	Student B
Need for Achievement	---	474
English Skills	470	566
Quantitative Ability	674	---

Note that a comparison based on the common dimension involves an evaluation of the difference between two scores on the same test, whereas a comparison based on the unique dimensions requires an evaluation of the relative contributions of two different tests. Because intradimensional comparisons are usually easier than interdimensional comparisons (Tversky, 1969; Russo & Dosher, 1983), the compatability hypothesis implies that the common dimension will be weighted more heavily than the unique dimension. This is precisely the effect observed by Slovic and MacPhillamy (1974). Interestingly, most subjects indicated, in a post experimental interview, that they did not intend to give more weight to the common dimension, and that they were unaware of doing so.

Another example of compatibility effects arises in studies of conceptual and perceptual similarity. Tversky and Gati (1978, 1982) showed that the relative weighting of common and distinctive features depends on their relation to the required task. More specifically, common features are weighted more heavily in judgments of similarity whereas distinctive features are weighted more heavily in judgments of dissimilarity. This effect produces reversals of order analogous to the reversals of preference. For example, familiar countries, such as East Germany and West Germany, were judged both more similar to each other and more dissimilar from each other than less familiar countries, such as Ceylon and Nepal. Indeed, the differential weighting scheme incorporated into the contrast model (Tversky, 1977) to describe such attentional shifts can be used to describe the effect of compatibility on preferences among gambles. The following section presents this model and shows how it applies to the pricing and rating data of the present study.

Differential Weighting Model

Let (p,x) be a gamble that offers a probability $0<p<1$ to win \$x, and probability $1-p$ to win nothing. Let $>_0$ be the choice order of gambles established by the comparison of bets, and let $>_1$ be the price order established by estimating worth. Suppose both orders satisfy a multiplicative model, as commonly assumed in theories of risky choice. Following prospect theory (Kahneman & Tversky, 1979) we use to denote the weighting of probabilities and v to describe the subjective value of monetary gains. Thus

$$(p,x) >_0 (q,y) \text{ iff } \pi(p)v(x) > \pi(q)v(y) \qquad (1)$$

Letting $f(p) = \log\pi(p)$ and $g(x)=\log v(x)$ we can express the model in an additive form.

$$\begin{aligned}(p,x) >_0 (q,y) &\text{ iff } f(p)+g(x) > f(q)+g(y)\\ &\text{ iff } g(x)-g(y) > f(q)-f(p).\end{aligned}$$

We assume that the price order is also additive but it gives more weight to the payoffs relative to the probabilities. That is,

$$\begin{aligned}(p,x) >_1 (q,y) &\text{ iff } f(p)+bg(x) > f(q)+bg(y) \qquad (2)\\ &\text{ iff } b[g(x)-g(y)] > f(q)-f(p)\end{aligned}$$

where $b>1$ is a bias parameter that reflects the accentuation of the payoffs induced by the pricing procedure. If $b=1$ the bias vanishes and the two orders coincide, whereas $b<1$ reflects a bias that amplifies probabilities relative to payoffs. Different elicitation procedures, or contexts, can be described by different values of b that express the relative contribution of probabilities and payoffs to the overall value of a gamble.

Using the common approximation that expresses the value of monetary increments as a power function $v(x)=x^a$, $a>0$, (Stevens, 1958; Tversky, 1967), (2) reduces to

$$(p,x)>_1(q,y) \text{ iff } \pi(p)x^{ab}>\pi(q)y^{ab} \qquad (3)$$

Hence, in a power utility model, the bias parameter b merely multiplies the exponent of the utility function. This transformation offers a simple way for incorporating a compatibility bias into prospect theory and other models that apply to more complicated gambles as well.

The differential weighting model defined in (1) and (2) was introduced as the simplest formal account of preference reversals, which requires only a single additional parameter for each response mode or preference order. The following discussion analyzes the qualitative assumption that underlies the model and provides it with an axiomatic basis.

Note that the proposed account of the compatibility effect leaves the scales f and g (or equivalently π and v) essentially unchanged; it merely modifies the slope of their indifference curves or the "rate of exchange" between probability and money. Consequently, the order of "probability intervals" or "monetary intervals" is preserved under different elicitation procedures, although the preference orders generated by these procedures do not coincide. That is, if the change from p to q has a bigger impact than the change from r to s, according to $>_0$, then the same conclusion must hold for $>_1$ as well. This condition, called <u>partial invariance</u>, can be restated in terms of the observed preferences $>_0$ and $>_1$ as follows. Suppose $w<x<y<z$ and $p<q<r<s$, then

$$(p,z)>_0(q,y),\ (q,w)>_0(p,x) \text{ and } (r,x)>_1(s,w)$$
$$\text{imply } (s,y)>_1(r,z) \qquad (4)$$

and the same relation holds when either the attributes or the orders are interchanged. A graphical illustration of this property, in which the inequalities are represented as arrows, is shown in Figure 1. Partial invariance is equivalent to the triple cancellation condition of additive conjoint measurement (see Krantz, Luce, Suppes & Tversky, 1971; Tversky, 1967), except that it applies here to the case of two order relations. The significance of partial invariance is that it is both necessary and sufficient for the differential weighting model, defined in (1) and (2).

Insert Figure 1 about here

<u>Theorem</u>

Let $>_0$ and $>_1$ be two additive order relations on the same set of gambles. That is,

$$(q,x)>_i(p,y) \text{ iff } f_i(q)+g_i(x)>f_i(p)+g_i(y),\ i=0,1.$$

Then the differential weighting model holds (i.e., $f_1=f_0$ and $g_1=bg_0$) if and only if partial invariance (4) is satisfied.

The proof of this theorem is given in the appendix.

The differential weighting model can be applied to the pricing and rating data from the present experiment. In Figure 2, each of the 12 gambles is plotted as a point in the probability x money plane on a double-log scale. The mean ratings and mean prices were regressed against the coordinates of the gambles. The multiple correlations were .98 for the prices and .96 for the ratings, indicating that these data are well approximated by a simple additive model. The slopes of the lines plotted in the figure give the ratios of the regression weights associated with the two coordinates. These slopes, which equal 2.70 for the ratings and .75 for the prices, reflect the tradeoff between probability and money for the two elicitation methods. Hence, the ratio of the two slopes 2.7/.75 = 3.6 provides an estimate of b, which is interpreted in this case as the degree to which the relative weight of probability to payoff is higher in rating than in pricing. Note that the above analysis is more restrictive than the general differential weighting model of Equation 2. It makes the additional assumptions that the subjective and the objective scales are related by power functions, yielding linearity on a logarithmic scale.

Insert Figure 2 about here

The ordering of the projections of the points on each of the two lines in Figure 2 (denoted by notches) represents the preference ordering induced by the two elicitation methods. Note that the projections of all the P bets (denoted by odd numbers) on the rating line exceed the projections of all the $ bets (denoted by even numbers). The negative correlation between the ratings and the prices (r = .-35 on the original scale, and r = -.30 on the log scale) is reflected by the numerous reversals of ordering of the projections on the two lines.

The differential weighting model could also be used to assess the relative importance of common vs. unique dimensions in the study by Slovic and MacPhillamy (1974). In that study, an additive model fit the comparisons between students quite well (R=.85). The mean weights for common and unique dimensions, respectively, were .83 and .73, producing a bias parameter of 1.15. The bias observed in the present study, therefore, is considerably stronger than that induced by the commonality of dimensions.

Discussion

Our findings show that strategically equivalent elicitation procedures give rise to markedly different preferences. We have attributed these results to the differential weighting of the components of the gamble, which is determined by the ease of mapping the components into the required response. Performance in perceptual-meter tasks has long been known to depend on the degree of compatability between the stimulus display and the required response (Fitts & Seeger, 1953, Wickens, 1984). The present results extend this concept to incorporate differential weighting of stimulus components in judgment and decision tasks.

Reversals of preferences can also be produced by the process of anchoring and insufficient adjustment. Although this factor does not appear to play an essential role in the present study, which did not provide explicit anchors, there is a great deal of evidence that anchoring has a powerful effect on both judgments and choices. For example, we asked a group of 72 Stanford undergraduates to state the amount of cash that is as desirable as a gamble that offered 1/6 chance to win $25 and 5/6 chance to win $2. Half the subjects were asked to take the low outcome ($2) as their initial estimate and then adjust it upwards until they reach a suitable cash equivalent. The other half of the subjects were asked to take the high outcome ($25) as their initial estimate and then adjust it downwards. The latter group produced significantly higher prices, with a median price of $10 as compared to a median price of $5 in the former group. The role of anchoring in preferences between risky prospects has also been discussed by Johnson & Schkade (1984), Lopes and Ekberg (1980), and Hershey and Schoemaker (1984).

The present results contribute to a growing body of literature that challenges traditional models of choice on the grounds that peoples' preferences are often ill-defined, unstable, and subject to framing and elicitation effects (see, e.g., Fischhoff, Slovic & Lichtenstein, 1980; March, 1978; Kahneman & Tversky, 1984). The frequent and persistent violations of invariance that have now been observed in many contexts indicate that the discrepancy between normative and descriptive theory is deeper and harder to bridge than is generally realized. The dependence of preference on the framing of decisions and the mode of elicitation raises both theoretical and practical questions for decision analysis. How should a choice be framed and what method of elicitation (choice, pricing, rating) should be used? How do we resolve the incoherence generated by the use of different frames and response modes? Descriptive studies of the resolution of incoherence (see, e.g., Lichtenstein & Slovic, 1971; Slovic & Tversky, 1974; Tversky & Kahneman, 1983) indicate that people often do not know how to reconcile their own inconsistencies. Indeed, in the absence of invariance, the problem of eliciting unbiased preferences and beliefs may elude a satisfactory solution.

References

Fischhoff, B., Slovic, P. & Lichtenstein, S. (1980). Knowing what you want: Measuring labile values. In T. Wallsten (Ed.), Cognitive processes in choice and decision behavior. Hillsdale, N. J.: Erlbaum.

Fishburn, P. C. (1983). Normative theories of decision making under risk and under uncertainty. Research paper, 75th Anniversary Colloquium Series, Harvard Business School, Cambridge, Massachusetts.

Fitts, P. M. & Seeger, C. M. (1953). S-R compatibility: Spatial characteristics of stimulus and response codes. Journal of Experimental Psychology, 46, 199-210.

Goldstein, W. M. (1982). Inconsistent assessments of preference: Effects of stimulus presentation method on the preference reversal phenomenon. Unpublished manuscript, Department of Psychology, University of Michigan.

Grether, D. M. & Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon. American Economic Review, 69, 623-638.

Hamm, R.M. (1984). The conditions of occurrence of the preference reversal phenomenon. Unpublished manuscript, Institute of Behavioral Science, University of Colorado.

Hershey, J. C. & Schoemaker, P. J. H. (1984). Probability vs. certainty equivalence methods in utility measurement: Are they equivalent?. Unpublished manuscript, University of Penn.

Johnson, E. & Schkade, D.A. (1984). Anchoring, adjustment and bias in utility assessments. Unpublished manuscript. Carnegie-Mellon University.

Kahneman, D. & Tversky, A. (1979). Prospect theory. Econometrica, 47, 263-292.

Kahneman, D. & Tversky, A. (1984). Choices, values, and frames. American Psychologist, 39, 341-350.

Krantz, D. H., Luce, R. D., Suppes, P. & Tversky, A. (1971). Foundations of Measurement, Vol. 1. New York: Academic Press.

Lichtenstein, S. & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. Journal of Experimental Psychology, 89, 46-55.

Lichtenstein, S. & Slovic, P. (1973). Response-induced reversals of preference in gambling: An extended replication in Las Vegas. Journal of Experimental Psychology, 101, 16-20.

Lindman, H R. (1971). Inconsistent preferences among gambles. Journal of Experimental Psychology, 89, 390-397.

Loomes, G. & Sugden, R. (1983). A rationale for preference reversal. American Economic Review, 73, 428-432.

Lopes, L.L. & Ekberg, Per-Hakan S. (1980). Test of an ordering hypothesis in risky decision making. Acta Psychologica, 45, 161-167.

March, J. G. (1978). Bounded rationality, ambiguity, and the engineering of choice. The Bell Journal of Economics, 9, 587-608.

Mowen, J.C., & Gentry, J.W. (1980). Investigation of the preference-reversal phenomenon in a new product introduction task. Journal of Applied Psychology, No. 6, 65, 715-722.

Pommerehne, W.W., Schneider, F., & Zweifel, P. (1982). Economic theory of choice and the preference reversal phenomenon: A reexamination. American Economic Review, 72, 569-574.

Reilly, R.J. (1982). Preference reversal: Further evidence and some suggested modifications in experimental design. American Economic Review, 72, 576-584.

Russo, J. E. & Dosher, B. A. (1983). Strategies for multiattribute binary choice. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9, 676-696.

Slovic, P., Fischhoff, B. & Lichtenstein, S. (1982). Response mode, framing, and information processing effects in risk assessment. In R. M. Hogarth (Ed.), New directions for methodology of social and behavioral science: The framing of questions and the consistency of response. San Francisco: Jossey-Bass.

Slovic, P. & Lichtenstein, S. (1968). The relative importance of probabilities and payoffs in risk-taking. Journal of Experimental Psychology Monograph Supplement, 78(3), part 2.

Slovic, P. & Lichtenstein, S. (1983). Preference reversals: A broader perspective. American Economic Review, 73, 596-605.

Slovic, P. & McPhillamy, D. J. (1974). Dimensional commensurability and cue utilization in comparative judgment. Organizational Behavior and Human Performance, 11, 172-194.

Tversky, A. (1967). Additivity, utility and subjective probability. Journal of Mathematical Psychology, 4, 175-202.

Tversky, A. (1969). Intransitivity of preferences. Psychological Review, 76, 31-48.

Tversky, A. (1977). Features of similarity. Psychological Review, 84, 327-352.

Tversky, A. & Gati, I. (1978). Studies of similarity. In E. Rosch & B. Lloyd (Eds.), Cognition and categorization. Hillsdale, N. J.: Erlbaum.

Tversky, A. & Gati, I. (1982). Representations of qualitative and quantitative dimensions. Journal of Experimental Psychology: Human Perception and Performance, 8, 325-340.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1131.

Tversky, A. & Kahneman, D. (1981). The framing of decisions and the psychology of choice. Science, 211, 453-458.

Tversky, A. & Kahneman, D. (1983). Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgment. Psychological Review, 90, 293-315.

Tversky, A. & Kahneman, D., (1984). Preference, invariance, and framing: Transparent and opaque choice. Unpublished manuscript, Stanford University.

Wickens, C. D. (1984). Engineering psychology and human performance. Columbus: Merrill.

Appendix

Theorem: Let $>_i, i = 0,\ldots,k$, be a family of preference relations on a set $P \times X$ of simple gambles satisfying

$$(p,y)>_i(q,x) \text{ iff } f_i(p)+g_i(y)>f_i(q)+g_i(x)$$

for all p,q in P and x,y in X. Then there exist functions f and g and constants b_i, such that $f_i = f$ and $g_i = b_i g$ if and only if partial invariance (4) holds. That is, $(p,z)>_i(q,y)$, $(q,w)>_i(p,x)$ and $(r,x)>_j(s,w)$ imply $(r,z)>_j(s,y)$, for all i and j.

Proof: To establish the necessity of partial invariance note that, by the differential weighting model,

$$(p,z)>_i(q,y) \text{ implies } f(p)+b_i g(z)>f(q)+b_i g(y)$$

$$(q,w)>_i(p,x) \text{ implies } f(q)+b_i g(w)>f(p)+b_i g(x).$$

Consequently, $g(z)-g(y)>g(x)-g(w)$. Furthermore

$(r,x)>_j(s,w)$ implies $f(r)+b_j g(x)>f(s)+b_j g(w)$. Thus, $b_j[g(x)-g(w)]>f(s)-f(r)$, and by the above inequality,

$b_j[g(z)-g(y)]>f(s)-f(r)$, and

$f(r)+b_j g(z)>f(s)+b_j g(y)$, hence

$(r,z)>_j(s,y)$ as required.

To prove sufficiency, note that under partial invariance, the inequalities

$$f_i(q)-f_i(p)>f_i(s)-f_i(r) \text{ and}$$

$$g_i(x)-g_i(w)>g_i(z)-g_i(y)$$

are independent of i, $0 \le i \le k$. Hence, there exist functions f and g and constants c_i, d_i such that for all $0 \leq i \leq k$,

$$(p,y) >_i (q,x) \text{ iff } c_i f(p) + d_i g(y) > c_i f(q) + d_i g(x).$$

If f_i and g_i are interval scales, the result follows immediately from the uniqueness of the scales; otherwise, one can construct f_i's (and g_i's) that are linearly related. Letting $b_i = d_i/(c_i+d_i)$ completes the proof of the theorem.

Acknowledgement

This work was sponsored by the Office of Naval Research under Grant NR 197-058 to Stanford University and by Contract N00014-82-C-0643 to Perceptronics, Inc. We wish to thank Maya Bar-Hillel, Sarah Lichtenstein and John Payne for comments on an earlier draft.

Table 1

Stimulus Gambles

Pair		P Bet		$ Bet
I.	1.	35/36 to win $4.00	2.	11/36 to win $16.00
II.	3.	29/36 to win $2.00	4.	7/36 to win $ 9.00
III.	5.	34/36 to win $3.00	6.	18/36 to win $ 6.50
IV.	7.	32/36 to win $4.00	8.	4/36 to win $40.00
V.	9.	34/36 to win $2.50	10.	14/36 to win $ 8.50
VI.	11.	33/36 to win $2.00	12.	18/36 to win $ 5.00

Table 2
Effect of Response Mode on Evaluations of Gambles

	Group A: Rating versus Choice	Rating P	Rating $		Group B: Price Versus Choice	Price P	Price $		Group C: Rating versus price	Rating P	Rating $	
Pair I	Choice P	69	3	.78	Choice P	20	18	.67	Price P	6	0	.19
	Choice $	16	4		Choice $	5	14		Price $	22	4	
		.92				.44				.88		
Pair II	Choice P	61	4	.73	Choice P	8	33	.75	Price P	3	1	.14
	Choice $	15	9		Choice $	2	12		Price $	21	4	
		.85				.18				.83		
Pair III	Choice P	59	1	.67	Choice P	6	28	.62	Price P	2	0	.07
	Choice $	26	4		Choice $	6	15		Price $	23	2	
		.94				.22				.92		
Pair IV	Choice P	59	1	.65	Choice P	17	20	.67	Price P	6	1	.22
	Choice $	23	10		Choice $	7	11		Price $	20	4	
		.88				.44				.87		
Pair V	Choice P	47	1	.54	Choice P	7	28	.62	Price P	3	0	.10
	Choice $	31	10		Choice $	0	21		Price $	22	5	
		.88				.12				.83		
Pair VI	Choice P	36	2	.46	Choice P	4	27	.56	Price P	1	0	.04
	Choice $	35	10		Choice $	1	23		Price $	20	4	
		.86				.09				.84		
Total	Choice P	331	12	.64	Choice P	62	154	.65	Price P	21	2	.13
	Choice $	146	47		Choice $	21	96		Price $	128	23	
		.89				.25				.86		

Note: Numbers of subjects exhibiting each response pattern are shown within boxes. Proportions of pairs in which the P Bet was evaluated more favorably than the $ Bet are shown in the margins.

Table 3

Effect of Response Mode on Mean Evaluations

of P Bets and $ Bets

Gamble	Expected Value	Mean Price	Mean Attractiveness Rating
35/36,4[a]	3.89	3.32	18.9
11/36,16	4.89	4.38	11.0
29/36,2	1.61	1.25	13.2
7/36,9	1.75	2.11	7.5
34/36,3	2.83	2.38	17.4
18/36,6.50	3.25	2.87	11.9
32/36,4	3.56	2.92	16.8
4/36,40	4.44	6.53	9.2
34/36,2.50	2.36	1.86	16.5
14/36,8.50	3.30	2.93	10.9
33/36,2	1.83	1.47	16.2
18/36,5	2.50	2.17	12.1
Overall P	2.68	2.20	16.5
Overall $	3.36	3.50	10.4

[a] Read: 35 chances out of 36 to win $4.00.

Table 4

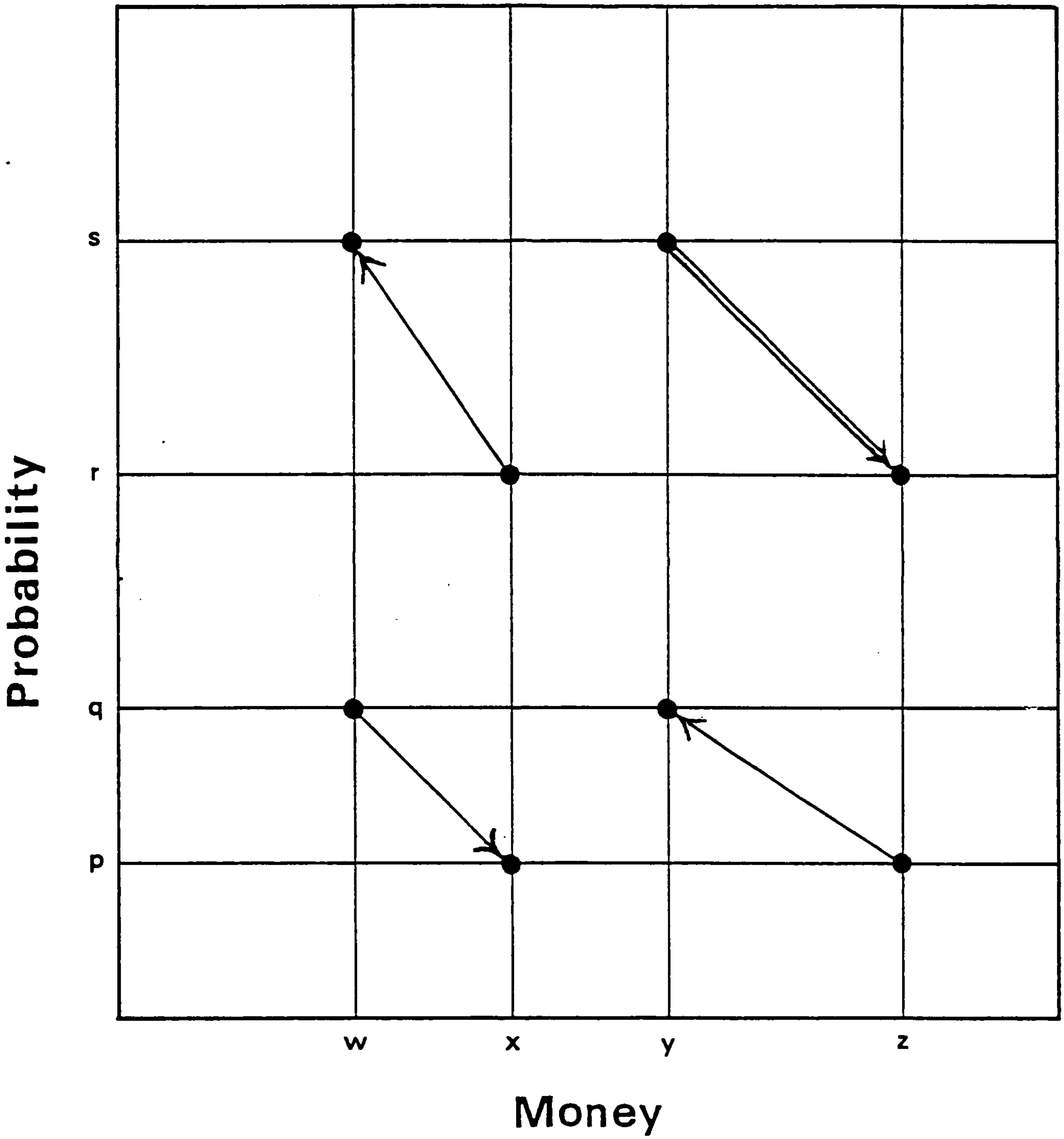
Test for Inflated Price Responses

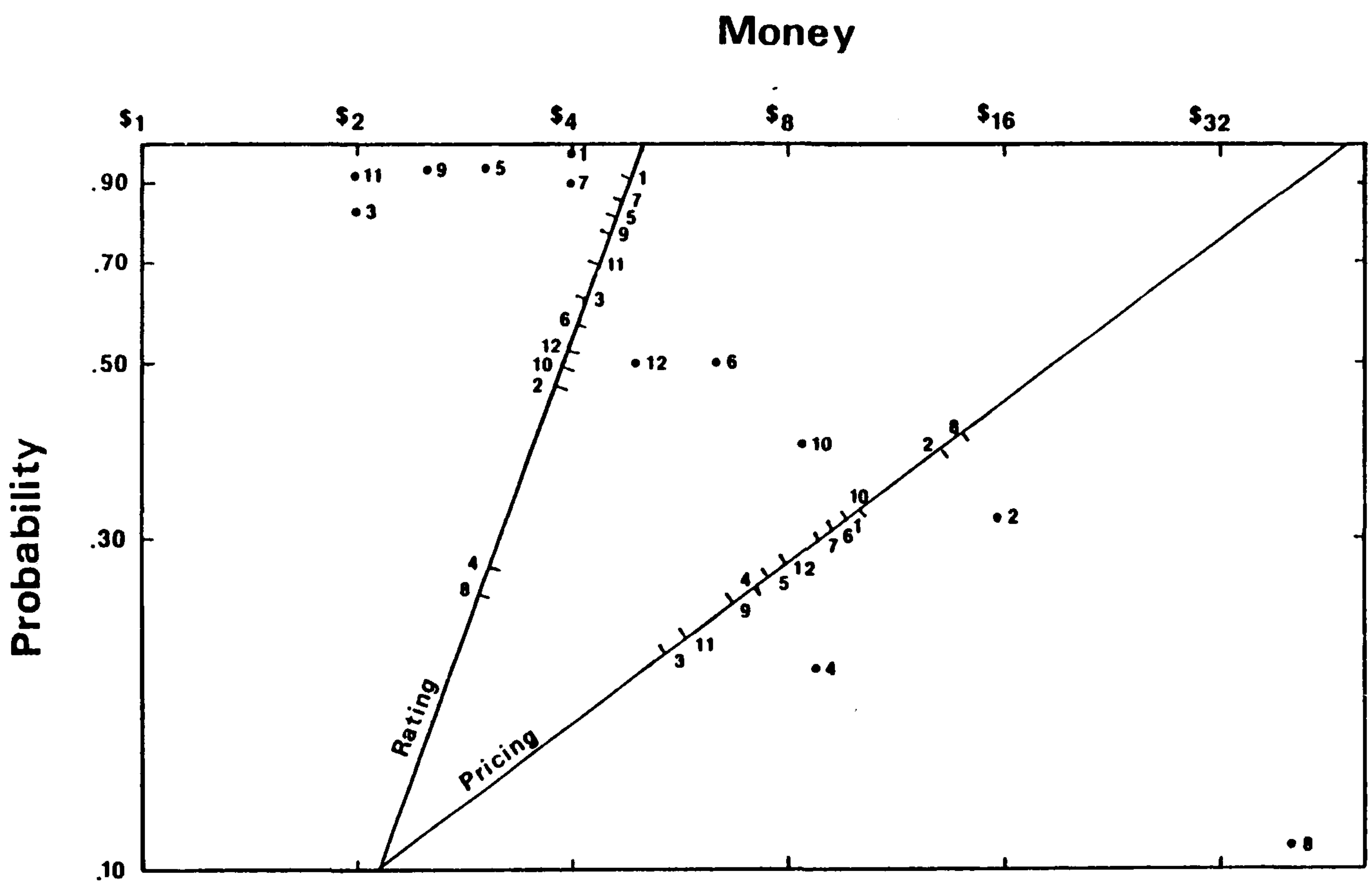
Bet		Sure Gain X	Percent of prices less than X	Percent choice of X over Bet
		P Bets		
1. 35/36,4	vs.	3.85	55	42
3. 29/36,2	vs.	1.50	66	59
5. 34/36,3	vs.	2.75	46	41
7. 32/36,4	vs.	3.25	48	47
9. 34/36,2.50	vs.	2.40	71	53
11. 33/36,2	vs.	1.85	63	62
	Overall		58	51
		$ Bets		
2. 11/36,16	vs.	5.75	78	86
4. 7/36,9	vs.	2.25	71	83
6. 18/36,6.50	vs.	3.25	71	66
8. 4/36,40	vs.	5.00	75	77
10. 14/36,8.50	vs.	3.00	56	72
12. 18/36,5	vs.	2.50	81	69
	Overall		72	75

Figure Captions

Figure 1. A graphical illustration of partial invariance. The hypotheses appear as arrows and the conclusion as a double arrow.

Figure 2. Best fit lines for pricing and rating data based on the differential weighting model. The slope of each line represents the tradeoff between probability and value for each elicitation method. The points represent the stimulus gambles. Their projections onto the best fit lines represent the predicted mean ratings and prices for each gamble under the model.


s
r
q
p
Probability
w
x
y
z
Money



Money
$1
$2
$4
$8
$16
$32
Probability
.90
.70
.50
.30
.10
Rating
Pricing

Risk Analysis, Vol. 3, No. 4, 1983

Judged Lethality: How Much People Seem to Know Depends Upon How They Are Asked

Baruch Fischhoff and Don MacGregor[1]

Received January 26, 1983; revised September 17, 1983

Four formally equivalent response modes were used to elicit laypeople's beliefs regarding the lethality of various potential causes of death. Results showed that respondents had an articulated core of beliefs about lethality that yielded similar orderings of maladies by lethality regardless of the response mode used. Moreover, this subjective ordering was fairly similar to that revealed by public health statistics. However, the absolute estimates of lethality produced by the different response modes varied enormously. Depending upon the mode used, respondents were seen to greatly overestimate or greatly underestimate lethality. The implications of these discrepancies for public education and risk analysis are explored.

KEY WORDS: Risk perception; judgment; risk assessment; elicitation.

1. INTRODUCTION

A recurrent question in the management of hazardous technologies is "How well does the public understand them?" Different answers can point to rather different roles for the public in hazard management. A well-informed public can be trusted to use technologies wisely, fend for itself in the marketplace, and identify its best interest in political decisions. An ignorant public may need protection from regulatory agencies, help to grasp political questions, or special training and safeguards to prevent misuse of potentially dangerous machines and substances.

At first blush, assessing the public's knowledge would seem quite straightforward. Just ask questions like: What is the probability of a nuclear-core meltdown? How many people die annually from asbestos-related diseases? and How does wearing a seat belt affect your probability of living through the year? The responses can be compared with the best available technical estimates, and deviations can be interpreted as showing the extent of the respondents' ignorance. This straightforward (jusk-ask-them) strategy is clearly superior to relying on speculation or anecdotal evidence.

There are, however, a number of constraints on it. A first constraint on questioning is that the questions address pertinent topics.[1] Laypeople have no way of knowing the answers to questions that concern classified, proprietary, or otherwise unpublished information. There is no reason (other than curiosity) for them to know facts that cannot affect their behavior. A second constraint is that the question be clear.[2,3] Jargon must be avoided, as must terms such as "risk," that seem clear but are used differently by different people.[4,5]

Our concern here is with a third constraint, one that remains even with questions that are worth asking and wording that is clear. It is the need to request knowledge in a form that is compatible with people's customary way of thinking about the topic. To acquit themselves properly in an interview, people must be able to express what they know. If the mental representation of their knowledge is different from the formulation required by the interviewer,

[1]Decision Research, A Branch of Perceptronics, 1201 Oak Street, Eugene, Oregon 97401.

then some translation is necessary, first to retrieve what they know and, second, to express what they retrieve. The greater the incompatibility, the more cumbersome the translation process becomes and the more knowledge is lost in transmission.

As a concrete example of possible difficulties, consider a group of (somewhat morbid) individuals who conscientiously read the obituaries in their local newspaper and have perfect recall. They are asked by an interviewer to estimate the relative frequency of different causes of death (or the age distribution of deaths) in their community. Although the respondents have all the requisite knowledge, in order to satisfy the interviewer they must aggregate it into the particular summary categories requested and perform the needed mental arithmetic in the time allotted.

One solution to the compatibility problem is *convergent validation*, eliciting judgments in several ways and trusting only patterns that emerge however the question is posed.[6] Although methodologically valid, convergent validation is a conservative strategy. It ignores many data and evades the compatibility problem by taking a position neither on how knowledge is represented in people's minds, nor on how best to extract it. A more direct approach is developed here within the specific context of eliciting judgments of the lethality of potential causes of death. This method builds upon convergent validation to identify core knowledge, which emerges however questions are posed. However, it also provides enough insight into the mental representation of knowledge to make some informed guesses about what method is best when discrepancies are observed.

2. THE STUDY

Although "risk" can be (and often is) spoken of as a uniquely defined, unitary concept, it clearly is not.[7] There are many different aspects of risk[8,9,10] and various ways to measure each.[11,12] One aspect of risk with an important influence on people's attitudes towards technological hazards is its degree of "lethality," the likelihood that if something goes wrong it will prove fatal.[5,8,13,14] All other things being equal, more lethal problems are viewed as more "risky" and in need of stricter regulation.

The present experiments consider lay estimates of the lethality in the U.S. of the 20 potential causes of death appearing in Table I. As a standard of comparison, the right-hand column offers statistical estimates derived from public health statistics. Although used as a standard, these statistics are not infallible. Poor sampling, incomplete reporting, and inconsistent attribution of multiply-caused deaths are some of the problems that make this a comparison between lay estimates and technical estimates (rather than between "real" and "perceived" risk).

The lay estimates here were elicited by four formally equivalent response modes; exemplary versions of which are:

(a) *Estimate death rate*: In a normal year, for each 100,000 people who have influenza, how many people do you think die of influenza?
(b) *Estimate number died*: Last year, 80,000,000 people had influenza. How many of them do you think died of it?
(c) *Estimate survival rate*: In a normal year, for each person who dies of influenza, how many do you think have influenza but do not die of it during the year?
(d) *Estimate number survived*: In a normal year, 5,000 people die of influenza. How many people do you think have influenza, but do not die from it during the year?

The formal equivalence of these four questions carries no assurance of their psychological equivalence. Each requires respondents to approach, translate, and express what they know in a somewhat different way. To the extent that the four questions elicit consistent estimates, one can conclude that respondents have a core of knowledge about lethality that is equally accessible from all four perspectives, and whose translation into a numerical response poses no problem. Conversely, inconsistent responses reveal the differential compatibility between response modes and knowledge representation.

Some potentially significant differences among the response modes are: (a) the death rate and survival rate conditions called for estimates of rates, whereas the number died and number survived conditions called for estimates of numbers; (b) those two conditions provided some information (which did not "give the answer away," but might have confirmed or contradicted existing beliefs); (c) the death rate and number died conditions dealt with fatalities, whereas the survival rate and number survived conditions dealt with survivors; (d) the correct answers for the number survived condition were generally much larger numbers than for the death rate, number died, and survival rate conditions (the medians were 3,000,000; 80; 5,500; and 1,250, respectively).

Table I. Direct and Converted Lethality Rate Estimates Based on Geometric Mean Responses

	Death rate per 100,000 afflicted				
Malady	Estimated death rate[a]	Estimated number died	Estimated survival rate	Estimated number survived	Statistical death rate
Dental problems	10	1	2	1	1
Influenza	393	6	26	511	6
Mumps	44	114	19	4	12
Skin diseases	63	4	6	641	30
Asthma	155	12	14	599	33
Alcoholism	559	70	13	294	44
Venereal disease	91	63	8	111	50
Measles	52	187	18	28	75
High blood pressure	535	89	17	538	76
Drug abuse	1,020	1,371	19	95	80
Bronchitis	162	19	43	2,111	85
Pregnancy	67	24	13	787	250
Diabetes	487	101	52	5,666	800
Emphysema	1,153	1,998	70	5,417	1,423
Tuberculosis	852	1,783	188	8,520	1,535
Pneumonia	563	304	77	9,553	1,733
Automobile accidents	6,195	3,272	31	6,813	2,500
Strokes	11,011	4,648	181	24,758	11,765
Heart attacks	13,011	3,666	131	27,477	16,250
Cancer	10,889	10,475	160	21,749	37,500
Coefficient of concordance	.62	.67	.34	.67	
N	40	38	40	40	

[a]Only these rates were estimated directly. Participants in other groups estimated other quantities, which were converted to lethality rates as described in the text.

3. EXPERIMENT 1

3.1. Method

One hundred and fifty-eight individuals were recruited through an advertisement in a university newspaper and paid for participating in this and several other unrelated studies of judgment and decision making. They were evenly divided between men (median age = 24) and women (median age = 21). The task was described in written instructions that provided some pertinent risk statistics, including the overall lethality rate for the U.S. (expressed in the terms of the ensuing questions). The 20 questions were then presented in a single randomized order.

All responses were converted to a common response mode, death rate per 100,000, to facilitate comparisons. Individual subjects' converted responses were summarized by geometric, rather than arithmetic, means so as to reduce the influence of outliers.

3.2. Results

The bottom row in Table I presents coefficients of concordance for each group. This statistic measures the degree of agreement among subjects within a group, with regard to the *ranking* of maladies by judged lethality. It ranges from 1.0 representing total agreement to 0.0 meaning lack of any agreement. As can be seen, there was fairly high agreement within the death rate, number died, and number survived groups, but rather low agreement within the survival rate group. This suggests that individuals from this population have fairly similar ideas regarding the relative lethality of these maladies, but that this consensus cannot express itself in the survival rate response mode.

The body of Table I presents the geometric means of the derived death rates. The four columns differ markedly in the magnitude of the numbers they include. These differences provide a clear ordering of the response modes by the magnitude of the estimates they produce, with number survived estimates being greatest followed by death rate and survival rate estimates. In extreme cases (e.g., cancer, strokes), estimates produced by the different methods range over two orders of magnitude. Despite these discrepancies in absolute estimates, there was general agreement regarding the relative lethality of these 20 maladies. Rank correlations between the entries in Table I ranged from .72 to .83 (all statistically significant; $p < .001$).

The similarity of the survival rate ordering to those of the other groups, despite the large differences in absolute values, is further evidence that this mode was incompatible with subjects' natural mode of thought. Expressing their core of knowledge in this form required a translation process that took much effort and added noise to subjects' judgments. That noise was reduced agreement among individuals, as seen in the low coefficient of concordance. However, such random errors cancelled out when subjects' responses were aggregated.

In a correlational sense, all response modes produced judgments that were closely related to the statistical estimates. Rank correlations between geometric mean estimates and the statistical estimates ranged from .82 (survival rate) to .86 (number survived). As Table I shows, however, these high correlations obscure substantial differences in the accuracy of the actual estimates. In general, the statistical death rates fell in the middle of the four sets of estimated rates. Thus, whether these individuals tended to over- or under-estimate lethality depends upon how the question was asked.

One measure of accuracy is an *error factor*, equal to the ratio of the estimated rate for a malady to the statistical rate, when the former is larger, or the reciprocal of that ratio, when the latter is larger. When computed over all individual responses, the geometric mean error factor for survival rate subjects was 33.2. By contrast, subjects in the other groups were, on the average, off by only a factor of 10 or so (see Table III, bottom).

4. EXPERIMENT 2

Apparently, people have a core of knowledge regarding relative lethality that emerges however they are queried. Moreover, the ordering roughly matches that provided by public health statistics. Both the magnitude and the reliability of their responses are, however, quite sensitive to the precise response mode used, with the survival rate question producing particularly low and unstable responses. Before interpreting these results in too great detail, it is worthwhile establishing how robust they are and clarifying the psychological processes involved in them. Experiment 2 attempts to do that by repeating and elaborating the tasks of Experiment 1.

4.1. Method

One hundred forty-three individuals repeated the tasks of Experiment 1 with a subset of 10 of the maladies for which public health statistics seemed most trustworthy—thereby allowing subjects to focus on few items. There were 37 subjects in the death rate group, 36 in the number died group, 37 in the survival rate group, and 36 in the number survived group. After answering, subjects were given the correct values for each item. In order to encourage attention to those values, they scored their own answers as too high or too low. After an hour of unrelated tasks, they were unexpectedly asked to recall the true value. Arguably, the best recall and the greatest improvement in knowledge will be with the most natural representation, that mode most conducive to the integration and preservation of additional knowledge. Finally, they saw the lethality of infectious hepatitis expressed in each of the four modes. They rated those phrasings by how "natural" they seemed, and how closely each "corresponds to the way you usually think of the lethality of diseases and accidents." An additional 87 subjects performed only this rating task.

4.2. Results

As shown in Table II, the initial estimates here resembled those from Experiment 1 (presented in Table I). Across the four groups, 26 of the 40 geometric mean estimates were within a factor of 2 of the comparable estimates from Experiment 1; all 40 were within a factor of 5. Again, the coefficients of concordance showed considerable agreement among subjects within each group except survival rate. Again, the overall orderings of the maladies within the different response modes were similar to one another and to the statistical estimates. Again, the statistical estimates fell below some group estimates and above others.

Table II. Initial and Recalled Lethality Rates: Experiment 2 (Geometric Means)

	Estimated death rate		Estimated number died		Estimated survival rate		Estimated number survived		Statistical rate[a]
	Initial	Recall	Initial	Recall	Initial	Recall	Initial	Recall	
Influenza	136	4	11	10	140	36	284	370	6
Asthma	59	49	12	35	33	397	858	115	30
Measles	57	57	401	407	67	321	61	37	75
Pregnancy	57	115	25	124	20	299	549	444	250
Diabetes	287	344	436	374	54	579	8,435	2,236	800
Emphysema	1,503	902	1,008	751	277	787	8,658	4,475	1,400
Tuberculosis	650	462	4,346	4,563	310	882	11,057	1,115	1,500
Pneumonia	482	352	392	156	199	854	9,279	9,580	1,700
Stroke	3,745	3,153	4,045	3,823	380	3,655	19,072	22,919	12,000
Cancer	6,110	12,106	9,211	8,433	327	7,388	17,526	33,128	37,000
Coefficient of concordance	.63	.58	.64	.66	.35	.33	.71	.80	
Rank correlation with statistical rate	.64	.87	.73	.64	.56	.78	.78	.78	

[a] Rates are given to subjects and are rounded to two significant figures.

After receiving the true values, subjects scored their own estimates as being too high or too low. One measure of the attention they paid is that there were only 47 errors in 1,480 scoring opportunities (= 3.2%).

The top section of Table III shows that in the unexpected recall task subjects infrequently remembered the statistical values that they had been given. The memory rate for individual maladies showed a serial position effect. The highest rates were for the first and last items (36.1%, influenza; 48.3%, pregnancy). The two worst remembered were fourth and sixth (0.6%, emphysema; 1.7%, tuberculosis). Personal relevance had some contribution to memorability insofar as cancer had the third best memory rate (22.4%) despite being fifth on the list. The second row of that table shows that when subjects failed to remember the correct value, they seldom supplanted it with their own initial estimate. Thus the two estimates were distinct enough in subjects' minds not to be confused.

The lower section of Table III shows that for all response modes, subjects' recollections were more accurate than their initial estimates. Thus, although subjects did not remember the statistical estimates, they did learn something from them. This learning was most pronounced with the survival rate group, whose recall estimates were, in the aggregate, as accurate as those of the other groups. Provision of the correct values seems to have enabled subjects to translate their ordinal knowledge of lethality into much more accurate numerical estimates for this response mode.

Eighty-seven "fresh" subjects rated the naturalness of the different modes for expressing information about the lethality of infectious hepatitis. Clearly, these *subjects thought it more natural* to think about lethality in terms of death than in terms of survival. There was no difference in preferences for statistic (rate or number). The rankings of the subjects who had previously completed the estimation and recall tasks were quite similar. Overall, mean rankings decreased by an average of 0.24 for subjects who had used a phrasing. Thus, although naturalness judgments are quite robust, they can be affected by immediate experience.

5. GENERAL DISCUSSION

In the aggregate, these results indicate that people have a fairly robust and consensual subjective ordering regarding the lethality of this set of maladies. The same ordering emerges with response modes sufficiently different to yield very different absolute estimates. This consistency means that it is possible to look at the substance of the lethality rankings regarding which maladies' relative lethality is overestimated or underestimated, although we will not do so.

Table III. Contrast between Original Estimates and Recall of True Values (Ten Items of Experiment 2)

	Estimated death rate	Estimated number died	Estimated survival rate	Estimated number survived
	Percentage of cases			
Recall = true value	19.3	15.3	21.3	10.4
Recall = initial estimate	4.1	4.4	3.0	5.2
	Geometric mean error factor			
Experiment 1 initial	10.9	10.2	33.2	12.5
Experiment 2 initial	12.5	10.7	43.0	12.1
Experiment 2 recall	4.2	6.5	7.6	9.2

That core of beliefs is not, however, as readily translated into all of the formally equivalent numerical expressions, as evidenced by differences in accuracy, within group agreement and naturalness ratings. The survival rate mode is clearly the outlier among these methods. It produced the least agreement among subjects and the worst absolute estimates. These results indicate a marked incompatibility between that response mode and subjects' customary ways of thinking about lethality. When respondents attempted to bridge that gap by themselves the result was noisy and biased responses. Along with number survived this mode was also rated least natural. Nonetheless, subjects were still able to exploit evidence presented in this mode, as shown by their vastly improved recall estimates. Thus, it appears harder to get information out of people with this mode than it is to get information into them.

Several simple accounts for these discrepancies in absolute judgments prove inadequate: (a) the *availability* explanation would argue that people are unduly influenced by the factors that are made most salient to them.[15] That should produce higher estimates of lethality with the response modes focused on death than with those focused on survival. However, the two survival response modes produced the largest and smallest lethality rates. (b) A *statistic* explanation would argue that the summary measure, a rate or numerical estimate, somehow affected performance. However, no such tendency was observed. (c) The same evidence would also reject a *storage mode* explanation: If people organize their information on a case-by-case basis, then the translation to a rate should be problematic; the converse would be true if subjects organized their knowledge in terms of rates. Yet, neither rates nor numbers were systematically higher or lower, more or less accurate, or more or less natural. (d) The number response modes provided some *additional information* (either the death toll or the affliction toll). In itself, that was not enough to improve performance consistently. (e) A *large number* explanation would argue that subjects have difficulty with response modes that require very large numbers,[16] which they are unaccustomed to using in daily life. For example, the number survived group was required to produce the largest numbers. Inability to do so would mean underestimating the number of survivors and emerge as overestimation of the lethality rate, the result obtained. The other groups, however, were required to produce numbers in a similar range, but showed quite different systematic biases. (f) An *anchoring and adjustment* explanation holds that respondents make quantitative estimates by picking some initially relevant number as a starting point and then adjusting it to accommodate additional information. In practice, that adjustment tends to be inadequate, turning the starting point into an anchor.[17] Unfortunately, the application of this heuristic with present tasks is unclear without independent knowledge of how people choose anchors. For example, was the number died group anchored on the total number of deaths, the number of deaths per 100,000 people in the U.S., the number of survivors, the number of deaths from accidents or from violent causes (all of which appeared on their form), or some other number(s) of their own creation?

Thus, none of these single factor explanations can account for the differences in the size of the magnitude estimates. Each might, of course, be "saved" if one could make an exception for one group or another. The most legitimate exception would be the survival rate group. If it is excluded, most of these explanations would prove quite serviceable, suggesting that each tells something about how people process such information.

5.1. Implications

The stable ordinal judgments observed here replicate the basic pattern observed in Lichtenstein's *et al.*[6] multi-method study of fatality judgments and Slovic's *et al.*[8] multi-method study of risk judgments. People have a consistent and fairly accurate feeling for the relative threat posed by different hazards. Where ordinal knowledge is all that is required, any response mode is good enough. However, if absolute estimates are needed, the methods matter greatly. People might respond quite differently to a threat if they assess its lethality by thinking about the survival rate or the number of survivors. A public health official could conclude that people underestimate or overestimate lethality, depending upon the question asked.

Our overall appraisal of the evidence produced by this multi-method approach suggests that the death rate and number died response modes provide the two best expressions of people's beliefs about lethality. They produce reliable and similar estimates; moreover, they are both judged to be quite natural. If this summary is correct, then it can be said that there is little systematic bias in people's lethality estimates.

We believe that some such multi-method analysis is essential before interpreting the responses produced with any response mode. The convergence found here is not assured. People might have had no coherent core of knowledge, knowing instead different things about death rates, survival rates, numbers died, and numbers survived. Responses to four such response modes would then tell four different stories. Assessing what people know would require evoking each perspective. Educators might be required to use several perspectives in order to ensure that people get the picture.

A needed extension of these methods is to the elicitation of information from technical experts in the context of risk analyses.[18, 19] For example, a supervisor might be asked how frequently workers fail to follow a particular operating procedure; an atmospheric chemist might be asked to assess a cumulative probability distribution for the oxidation rate in some complex situation; a mechanical engineer might be asked to estimate the failure rate for a familiar valve in an unfamiliar use. Such questions may be formulated for the convenience of the consumer of that knowledge (the risk analyst) or its producer (the technical expert). However, being an expert in a topic need not mean being an expert in answering questions about it. In that case, all formally equivalent questions are not psychologically equivalent. Question design may be as important an aspect of risk analysis as system modeling.

ACKNOWLEDGMENTS

This research was partially supported by the Advanced Research Project Agency of the Department of Defense under Contract N00014-79-312-0730 (ARPA Order No. 0194) under subcontract from Decisions and Designs, Inc. and by the National Science Foundation under Grant SES 8213452 to Perceptronics, Inc. Any opinions, findings and conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

1. B. Fischhoff, P. Slovic, and S. Lichtenstein, "The 'Public' vs. the 'Experts': Perceived vs. Actual Disagreements About the Risks of Nuclear Power," in V. Covello, G. Flamm, J. Rodericks, and R. Tardiff (eds.), *Analysis of Actual vs. Perceived Risks* (Plenum Press, New York, 1983).
2. National Academy of Sciences, *Survey Measure of Subjective Phenomena* (NAS, Washington, D. C., 1981).
3. S. L. Payne, *The Art of Asking Questions* (Princeton University Press, Princeton, New Jersey, 1980).
4. B. Fischhoff, P. Slovic, and S. Lichtenstein, "Lay Foibles and Expert Fables in Judgments about Risk," *American Statistician* **36**, 240–255 (1982).
5. P. Slovic, B. Fischhoff, and S. Lichtenstein, "Rating the Risks," *Environment* **21**, 14–20, 36–39 (1979).
6. S. Lichtenstein, P. Slovic, B. Fischhoff, M. Layman, and B. Combs, "Judged Frequency of Lethal Events," *Journal of Experimental Psychology: Human Learning and Memory* **4**, 551–578 (1978).
7. B. Fischhoff, S. Lichtenstein, P. Slovic, S. L. Derby, and R. L. Keeney, *Acceptable Risk* (Cambridge University Press, New York, 1981).
8. P. Slovic, B. Fischhoff, and S. Lichtenstein, "Characterizing Perceived Risk," in R. W. Kates and C. Hohenemser (eds.), *Technological Hazard Management* (Oelgeschlager, Gunn & Hain, Cambridge, Massachusetts, 1983).
9. C. Vlek and P. J. Stallen, "Judging Risks and Benefits in the Small and in the Large," *Organizational Behavior and Human Performance* **28**, 235–271 (1981).
10. D. von Winterfeldt, R. S. John, and K. Borcherding, "Cognitive Components of Risk Ratings," *Risk Analysis* **1**, 277–288 (1981).
11. E. A. C. Crouch and R. Wilson, *Risk/Benefit Analysis* (Ballinger, Cambridge, Massachusetts, 1982).

12. S. R. Watson, "On Risks and Acceptability," *Journal of the Radiological Protection Society* **1**, 21–25 (1981).
13. B. Fischhoff, P. Slovic, S. Lichtenstein, S. Read, and B. Combs, "How Safe Is Safe Enough? A Psychometric Study of Attitudes Towards Technological Risks and Benefits," *Policy Sciences* **8**, 127–152 (1978).
14. C. H. Green, "Risk: Attitudes and Beliefs," in D. V. Canter (ed.), *Behaviour in Fires* (Wiley, Chichester, 1982).
15. A. Tversky and D. Kahneman, "Availability: A Heuristic for Judging Frequency and Probability," *Cognitive Psychology* **4**, 207–232 (1973).
16. E. C. Poulton, "Biases in Quantitative Judgments," *Applied Ergonomics* **13**, 31–42 (1982).
17. A. Tversky and D. Kahneman, "Judgment under Uncertainty: Heuristics and Biases," *Science* **185**, 1124–1131 (1974).
18. M. G. Morgan, M. Henrion, and S. C. Morris, *Expert Judgment for Policy Analysis* (BNL 51358) (Brookhaven National Laboratory, Upton, New York, 1980).
19. U. S. Nuclear Regulatory Commission, *Fault Tree Guide* (NRC, Washington, D. C., 1981).

Behavioral Decision Theory Perspectives on Risk and Safety

Paul Slovic, Baruch Fischhoff, and Sarah Lichtenstein

Decision Research
A Branch of Perceptronics
1201 Oak Street
Eugene, Oregon 97401
USA

This work was supported by the Office of Naval Research under Contract N00014-82-C-0643 to Perceptronics, Inc.

Paper presented at the Conference on
Subjective Probability, Utility, and Decision Making
Groningen, The Netherlands, August 1983

Abstract

Behavioral decision theory can contribute in many ways to the management and regulation of risk. In recent years, empirical and theoretical research on decision making under risk has produced a body of knowledge that should be of value to those who seek to understand and improve societal decisions. This paper describes several components of this research, which is guided by the assumption that all those involved with high-risk technologies as promoters, regulators, politicians, or citizens need to understand how they and the others think about risk. Without such understanding, well-intended policies may be ineffective, perhaps even counterproductive.

Behavioral Decision Theory Perspectives on Risk and Safety

Paul Slovic, Baruch Fischhoff, and Sarah Lichtenstein

Decision Research, A Branch of Perceptronics

In modern industrial societies, the control of technological hazards has become a major concern of the public and a growing responsibility of government. Yet despite massive efforts to manage these hazards, many people feel increasingly vulnerable to their risks and believe that the worst is yet to come. Risk management agencies have become embroiled in rancorous conflicts, caught between a fearful and unsatisfied public on one side and frustrated technologists and industrialists on the other. The way in which these conflicts are resolved may affect not just the fate of particular technologies, but the fate of industrial societies and their social organization as well.

Research within the framework of behavioral decision theory can contribute in many ways to the management and regulation of risk. This paper describes several components of this research and its application to such practical problems as developing safety standards for hazardous technologies and creating programs to inform people about risk.

Informing People about Risk

One consequence of the growing concern about hazards has been pressure on the promoters and regulators of hazardous enterprises to inform citizens, patients, and workers about the risks they face from their daily activities, their medical treatments, and their jobs. Attempts to implement information programs depend upon a variety of political, economic and legal forces (e.g., Gibson, in press; Sales, 1982). The success of such efforts depends, in part, upon how clearly the information can be presented (Fischhoff, in press-a; Slovic, Fischhoff & Lichtenstein, 1980-a, 1981-a).

One thing that past research demonstrates clearly is the difficulty of creating effective risk-information programs. Doing an adequate job means finding cogent ways of presenting complex technical material that is clouded by uncertainty and may be distorted by the listeners' preconceptions of the hazard and its consequences. Difficulties in putting risks into perspective or resolving the conflicts posed by life's gambles may cause risk information to frighten and frustrate people, rather than aid their decision making.

If an individual has formed strong initial impressions about a hazard, results from cognitive social psychology suggest that those beliefs may structure the way that subsequent evidence is interpreted. New evidence will appear reliable and informative if it is consistent with one's initial belief; contrary evidence may be dismissed as unreliable, erroneous, or unrepresentative. As a result, strongly held views will be extraordinarily difficult to change by informational presentations (Nisbett & Ross, 1980).

When people lack strong prior opinions about a hazard, the opposite situation exists—they are at the mercy of the way that the information is presented. Subtle changes in the way that risks are expressed can have a major impact on perceptions and decisions. One dramatic recent example of this comes from a study by McNeil, Pauker, Sox, and Tversky (1982), who asked people to imagine that they had lung cancer and had to choose between two therapies, surgery or radiation. The two therapies were described in some detail. Then, some subjects were presented with the cumulative probabilities of surviving for varying lengths of time after the treatment. Other subjects received the same cumulative probabilities framed in terms of dying rather than surviving (e.g., instead of being told that 68% of those having surgery will have survived after one year, they were told that 32% will have died). Framing the statistics in terms of dying dropped the percentage of subjects choosing radiation therapy over surgery from 44% to 18%. The effect was as strong for physicians as for laypersons.

A rather different kind of effect may be seen in Table 1 which shows the results of asking people to estimate the chances of dying from various maladies, given that one had been afflicted with them. The first four columns show mean responses to four formulations of the question that are equivalent formally, but apparently quite different psychologically. Once converted to a common unit (deaths per 100,000), these response modes produce estimates differing greatly in magnitude. If these estimates were used as guides to policy making, then the respondents might seem to overestimate or underestimate the risks, depending upon what question they were asked. Conversely, presenting actuarially accurate information might have a quite different impact depending upon the formulation used.

Insert Table 1 about here

Numerous other examples of "framing effects" have been demonstrated by Tversky and Kahneman (1981) and Slovic, Fischhoff, and Lichtenstein (1982-a). Some of these effects can be explained in terms of the nonlinear probability and value functions proposed by Kahneman and Tversky (1979) in their theory of risky choice. Others can be explained in terms of other information-processing considerations such as compatibility effects, anchoring processes, and choice heuristics. Whatever the causes, the fact that subtle differences in how risks are presented can have such marked effects suggests that those responsible for information programs have considerable ability to manipulate perceptions and behavior.

The stakes in risk problems are high—industrial profits, jobs, energy costs, willingness of patients to accept treatments, public safety and health, etc. When subtle aspects of how (or what) information is presented can significantly change people's responses, the choice of formulation involves issues of law, ethics, and politics as well as behavioral decision theory.

One thing that behavioral research can offer to these decisions is an assessment of how large these effects are. When they are large, as in the examples given, the conflicts of interest may be so great that no one group can be entrusted with preparing informational statements. A second kind of guidance is describing the potential kinds of bias so that the parties involved can defend their own interests. A third contribution is assessing the feasibility of informational programs, that is, how well people can be informed. Fortunately, despite the evidence of difficulties, there is also evidence showing that properly designed information programs can be beneficial. Research indicates that people can understand some aspects of risk quite well and they do learn from experience. For example, even in Table 1, the orderings of risk judgments with the different response modes was highly consistent. In situations where misperception of risks is widespread, people's errors can often be traced to inadequate information and biased experiences, which educational programs may be able to counter. A final contribution is determining how interested people are in having the information at all. Despite occasional claims to the contrary by creators of risk, people seem to want all the information that they can get (Fischhoff, in press-b; Slovic, Fischhoff & Lichtenstein, 1980-a).

Characterizing Perceived Risk

One objective of research on risk perception has been to develop a taxonomy for hazards that could be used to understand and predict the way that society responds to them. Such a taxonomy might explain, for example, people's extreme aversion to some hazards, their indifference to others, and the discrepancies between these reactions and experts' views. During recent years, we and others have continued to employ what might be called the "psychometric paradigm," exploring the ability of psychophysical scaling methods and multivariate analysis techniques to produce meaningful quantitative representations of risk attitudes and perceptions (see, for example, Brown & Green, 1980; Gardner et al., 1982; Green, 1980; Green & Brown, 1980; Johnson & Tversky, in press; Lindell & Earle, 1982; MacGill, 1982; Renn, 1981; Slovic, Fischhoff & Lichtenstein, 1980b, in press; Vlek & Stallen, 1979; von Winterfeldt, John & Borcherding, 1981). Although each new study adds richness to the picture, some broad generalizations seem to be emerging.

Researchers exploring the psychometric paradigm have typically asked people to judge the current riskiness (or safety) of diverse sets of hazardous activities, substances, and technologies, and to indicate their desires for risk reduction and regulation of these hazards. These global judgments have then been related to judgments about: (i) the hazard's status on various qualitative characteristics of risk (e.g., voluntariness, dread, knowledge, controllability), (ii) the benefits that it provides to society, (iii) the number of deaths it causes in an average year, and (iv) the number of deaths it can cause in a disastrous accident or year.

Among the generalizations that have been drawn from the results of the early studies in this area are the following:

(1) Perceived risk is quantifiable and predictable. Psychometric techniques seem well suited for identifying similarities and differences among groups with regard to risk perceptions and attitudes.

(2) "Risk" means different things to different people. When experts judge risk, their responses correlate highly with technical estimates of annual fatalities. Laypeople can assess annual fatalities if they are asked to (and produce estimates somewhat like the technical estimates). However, their judgments of risk are sensitive to other factors as well (e.g., catastrophic potential, threat to future generations) and, as a result, may differ from their own (or experts') estimates of annual fatalities.

(3) Even when groups disagree about the overall riskiness of specific hazards, they show remarkable agreement when rating those hazards on characteristics of risk such as knowledge, controllability, dread, catastrophic potential, etc.

Most psychometric studies have been based on correlations among mean ratings of risk and risk characteristics across different technologies. If robust, the relationships revealed this way should be indicative of how society as a whole responds to hazards. They may also reflect the perceptions of most individuals looking at a set of hazards. However, as pointed out by Gardner et al. (1982) and Renn (1981), such relationships need not hold true at the level of individual respondents evaluating a single technology. For example, just because technologies judged to be relatively high in catastrophic potential also tend to be judged as high in risk does not mean that those persons who see a specific technology as particularly catastrophic will also perceive it as relatively risky. Understanding the relationships at this level help explain why certain individuals exhibit a high degree of concern about a particular technology. Some studies of this type are currently underway.

Factor Analytic Representations

Many of the qualitative risk characteristics are highly correlated with each other, across a wide domain of hazards. For example, hazards rated as "voluntary" tend also to be rated as "controllable" and "well known"; hazards that threaten future generations tend also to be seen as having catastrophic potential, etc. Investigation of these interrelationships by means of factor analysis has shown that the broader domain of characteristics can be condensed to two or three higher-order characteristics or factors.

The factor space presented in Figure 1 has been consistently replicated across groups of laypersons and experts judging large and diverse sets of hazards. The factors in this space reflect the degree to which a risk is understood, the degree to which it evokes a feeling of dread, and the number of people exposed to the risk. Making the set of hazards more specific (e.g., partitioning nuclear power into radioactive waste transport, uranium mining, nuclear reactor accidents, etc.) appears to have little effect on the factor structure or its relationship to risk perceptions (Slovic, Fischhoff & Lichtenstein, in press).

Insert Figure 1 about here

We have found that laypeople's risk perceptions and attitudes are closely related to the position of a hazard within the factor space. Most important is the factor "Dread Risk." The higher a hazard's score on this factor, the higher its perceived risk, the more people want to see its current risks reduced, and the more they want to see strict regulation employed to achieve the desired reduction in risk (Figure 2). Recently, we have also found that the informativeness or "signal potential" of an accident or mishap, which appears to be a key determiner of its social impact, is systematically related to both Dread Risk and Unknown Risk factors (see Figure 3).

Insert Figures 2 and 3 about here

Other Representations

The picture that has emerged from our factor analytic studies of perceived risk has been so consistent that one is tempted to believe in its universality. However, any such beliefs must be tempered in the face of recent evidence provided by other researchers.

Similarity-based representations. Factor-analytic studies supply respondents with the component characteristics of risk. An alternative approach is to have people rate the similarity of hazard pairs with regard to risk and to use some form of multidimensional scaling technique to construct a dimensional representation of the similarity space. Multi-dimensional scaling of similarity judgments for small sets of hazards by Vlek and Stallen (1979) and Green and Brown (1980) has produced two-dimensional representations similar to those obtained in our factor-analytic studies. However, Vlek and Stallen found substantial individual differences in the weighting of the dimensions.

Johnson and Tversky (in press) have compared factor analytic and similarity representations derived from the same set of 18 hazards. The hazards differed from those in Figure 1 in that they included natural hazards and diseases as well as activities and technologies. They found that the factor space derived from this different set of hazards was not quite the same as the space derived from our studies. Furthermore, they found that judgments of similarity based on direct comparisons of hazards were very different from similarity indices derived from evaluations of the hazards on a set of characteristics supplied by the experimenter. For example, homicide was judged to be similar to other acts of violence (war, terrorism) despite having a very different profile on the various risk characteristics.

In addition to producing a multidimensional representation of the similarity data, Johnson and Tversky constructed a tree representation (Figure 4). The risks are the terminal nodes of the tree and the distance between any pair of risks is given by the length of the horizontal parts of the shortest path that joins them; the vertical part is included only for graphical convenience. A tree representation can be interpreted in terms of common and unique features. Figure 4 exhibits a distinct hierarchy of clusters which Johnson and Tversky call: hazards, accidents, violent acts, technological disasters and diseases.

Insert Figure 4 about here

<u>The repertory grid</u>. Another way to derive risk characteristics is with the repertory grid technique. Green and Brown (1980) used this technique to generate data which were then analyzed by Perusse (1980). Subsets of three hazards were selected from a larger set of 21. Respondents were asked to indicate in what way two of the hazards are similar to each other and different from the third. The universe of constructs generated by this technique is shown in Table 2. Obviously, it includes many characteristics not studied previously. It would seem worthwhile to use the repertory grid as a starting point for factor analytic studies. In principle, each new item might be a predictor of people's behavior.

Insert Table 2 about here

<u>Free-response questionnaires</u>. The repertory grid can be viewed as a member of a larger class of free-response techniques, which allow respondents to generate their own response alternatives. Earle and Lindell (in press) have used such techniques to survey public perceptions of hazardous industrial facilities. Although many of their results replicate those from studies using structured response alternatives, they found some potentially important new findings. One was that their respondents failed to exhibit concern for future generations, in contrast to the concern shown in factor analytic studies and in all moral treatments of this topic.

The importance of these studies lies in what they reveal about the variation of hazard perception across tasks, item sets, and methods of analysis. If these differences prove to be reliable, then great care will be needed to choose the method most suitable to the purposes of particular research projects. As indicated above, factor analytic representations predict certain important attitudes towards hazards. Johnson and Tversky (in press) hypothesized that similarity-based representations may predict other responses, such as reactions to new risks or new evidence about risks (e.g., the effect of Tylenol poisoning on the purchase of over-the-counter drugs). The purpose is also important for the design of the experiment. Factor analyses conducted on diverse sets of items may miss "local" features pertinent to only a few hazards. Similarity judgments allow consideration of features that experimenters may have missed. However, similarity may be influenced by superficial or irrelevant considerations (e.g., electric power and nuclear power may be judged "similar" in "risk" because they are both sources of power).

<u>Implications of Risk Perception Research</u>

The social implications of the research we have been describing have been a matter of lively debate, taking up most of the June, 1982 issue of the journal, <u>Risk Analysis</u>. Douglas and Wildavsky (1982) have argued that psychometric studies, with their cognitive emphasis, omit social and cultural processes that play a major role in determining which risks society fears and which it ignores. Otway and Thomas (1982) have taken a particularly cynical view, arguing that this research is being used as a tool in a discourse which is not concerned with risks per se, nor with perceptual and cognitive processes. Rather, the hidden agenda is the legitimacy of decision-making institutions and the equitable distribution of hazards and benefits.

Our view (Slovic, Fischhoff & Lichtenstein, 1982-b) is that an understanding of how people think about risk has an important role in informing policy, even if it cannot resolve all questions. Moreover, risk perception research can be used to challenge social-political assumptions as well as to reinforce them (e.g., Fischhoff, Slovic & Lichtenstein, in press). Behavioral studies of flood-insurance decisions and seat-belt usage have already provided policy relevant insights. The psychometric studies described above provide the beginnings of a psychological classification system for hazards that may help explain and forecast reactions to specific technologies such as nuclear power or genetic engineering (see, e.g., Slovic, Lichtenstein, Fischhoff, in press) or provide guidelines for managing the social conflicts surrounding hazardous technologies (von Winterfeldt & Edwards, 1983).

One important contribution of existing research has been to demonstrate the inadequacy of the unidimensional indices (e.g., annual probability of death, loss of life expectancy) that have often been advocated for "putting risks in perspective" and aiding decision making. Psychometric studies suggest that such comparisons will be unsatisfactory because people's perceptions are determined not only by mortality statistics but also by a variety of quantitative and qualitative characteristics. These include a hazard's degree of controllability, the dread it evokes, its catastrophic potential, and the equity of its risk/benefit distribution. Attempts to characterize, compare, and regulate risks must be sensitive to the broader conception of risk that underlies people's concerns. Fischhoff, Watson, and Hope (1983) have made a start in this direction by demonstrating how one might go about constructing a more adequate definition of risk. They show that variations in the scope of one's definition of risk can greatly change the assessment of risk from various energy technologies.

The Search for Acceptable Risk

The third topic in our survey deals with the elusive search for an answer to the question, "How safe is safe enough?" The question takes such forms as: "Do we need improved emergency cooling systems in our nuclear power plants?" "Is the carcinogenicity of saccharin sufficiently low to allow its use?" "Should schools with asbestos ceilings be closed?"

Frustration over the difficulty of answering such questions has led to a search for clear, implementable rules that will determine whether a given technology is sufficiently safe, i.e., are its risks acceptable. Despite heroic efforts on the part of many risk analysts, no magic formula has been discovered. Nonetheless, some progress has been made, not the least of which includes a heightened respect for the complexities of the task.

Approaches to Acceptable Risk: A Critique

Our own efforts in this area during recent years have been instigated and supported by the Nuclear Regulatory Commission (NRC). It has always been known that nuclear reactors could be made safer—at increased cost. However, as long as it was difficult to quantify safety, the question of how much safety at what price was rarely addressed explicitly. The technology of measuring risk has advanced rapidly in recent years. Now that quantitative estimates of accident probabilities are thought to be accessible, the need to determine how safe reactors <u>should</u> be has taken on greater significance.

At the urging of Congress and the nuclear industry, the NRC has been working intensively to develop an explicit, possibly quantitative, safety goal or philosophy. Presumably this goal would clarify the Commission's vague mandate to "avoid undue risk to public health and safety" and would serve to guide specific regulatory decisions.

The NRC asked us to take a comprehensive, critical look at the philosophical, sociopolitical, institutional, and methodological issues crucial to answer the question of "How safe is safe enough?". We approached this task in a general way, not restricted to nuclear power or any other specific technology. Guided by behavioral decision theory, our examination of approaches to acceptable risk attempted to:

(a) Characterize the essential features of acceptable-risk problems that make their resolution so difficult. These features included uncertainty about how to define acceptable-risk problems, difficulties in obtaining crucial facts, difficulties in assessing social values, unpredictable human responses to hazards, and problems of assessing the adequacy of decision-making processes.

(b) Create a taxonomy of decision-making methods, described according to how they address the essential features of acceptable-risk problems. The major approaches we discussed were <u>professional judgment</u>: allowing technical experts to devise solutions; <u>bootstrapping</u>: searching for historical precedents to guide future decisions; and <u>formal analysis</u>: theorybased procedures for modeling problems and calculating the best decision, such as risk/benefit, cost/benefit, and decision analysis.

(c) Specify the objectives that an approach should satisfy in order to guide social policy. These included comprehensiveness, logical soundness, practicality, openness to evaluation, political acceptability, institutional compatibility, and conduciveness to learning.

(d) Evaluate the success of the approaches in meeting these objectives.

(e) Derive recommendations for policy makers and citizens interested in improving the quality of acceptable-risk decisions.

Space permits only a brief glimpse at our conclusions. Details can be found in Fischhoff, Lichtenstein, Slovic, Derby and Keeney (1981). Perhaps most important was the conclusion that acceptable-risk problems are decision problems, requiring a choice among alternatives. That choice depends on the set of options, consequences, values, and facts invoked in the decision-making process. Therefore, there can be no single, allpurpose number that expresses the acceptable risk for a society. At best, one can hope to find the most acceptable alternative in a specific problem. Indeed, "acceptable risk" may be a poor term if it connotes universality. Otway and von Winterfeldt (1982) have put forth a similar view, arguing in addition, that many non-risk factors must also be weighed in determining the acceptability of a technology.

We also concluded that each approach to acceptable risk was incomplete and biased, that separation of facts from values was desirable though usually infeasible, and that the way the problem is defined is often the determining factor in acceptable-risk decisions. Finally, the choice of a method for decision making should be recognized as a political issue, affecting the distribution of power and expertise within a society.

Toward a Safety Goal

Justification. Our analysis of decision-making approaches was used by the NRC in the planning stages of its program to develop a safety goal, stating how safe nuclear power must be. Upon completion of this analysis, we were asked to participate in the development of the goal itself. Before doing so, we felt it necessary to critique the effort in light of our earlier conclusion that, since acceptable risk is the outcome of specific decisions, there can be no single, all purpose number (standard or goal) that does the job. Beyond the obvious efficiency of setting a generally applicable decision rule, are there any other justifications for goals and standards? Fischhoff (1983, in press-c) wrestled with this question and concluded that there were, indeed, circumstances in which standards were warranted. Table 3 gives a list of conditions, any one of which might justify the development of a pass/no pass safety standard:

In addition to providing a theoretical rationale for goals and standards, these analyses explore the many subtle and complex problems involved in transforming a goal from a political statement to a useful tool, one that can be unambiguously applied by regulators and understood by the regulated. Here one faces issues such as (a) defining the category governed by the standard (e.g., Is a cosmetic a drug?); (b) determining the point and time of regulation (e.g., plant by plant or company by company? At which stage of production and use?); (c) tailoring standards to mesh with engineering and design capabilities; (d) deciding whether to regulate technical matters (nuts and bolts) or performance ("as long as you meet this goal, we don't care how you do it"). Once one has decided where to place the standard, a critical question involves how to measure risks in order to determine whether they are in compliance with the standard.

Social and behavioral issues. Having satisfied ourselves that general goals and standards had a place in the regulator's armamentarium (the NRC had presumed this), we proceeded to consider the detailed process of establishing a safety goal. Our objective was to critique, from our perspective as behavioral decision theorists, what tended to be seen as primarily a technical problem, dealing with the design, construction, and licensing of reactors and the ability of probabilistic techniques to assess and verify reactor risks.

There has been no shortage of proposed safety goals over the years. Solomon, Nelson, and Salem (1981) counted 103 criteria pertaining to reactor accidents, which they categorized as follows:

1. Criteria for the safety of reactor systems: e.g., an upper limit for the acceptable probability of a coremelt accident.

2. Differential criteria for the allowable risks to individuals in the vicinity of the plant site and distant from the plant site.

3. Criteria for the maximum allowable expenditures to avert a personrem of radiation exposure.

The criteria proposed by the NRC fell within these generic categories. A detailed discussion of these various criteria is beyond the scope of this paper. Suffice it to say that (a) they tend to be derived on the basis of comparisons with other accident risks and with the risks from other sources of electricity, (b) they are concerned with a rather narrow view of the costs of a reactor accident, focusing on immediate and latent fatalities, physical damage to the reactor and adjoining property, and costs of cleanup and replacement electricity, and (c) they sometimes incorporate risk aversion in the form of a weighting factor that attributes extra significance and cost to accidents that cause multiple fatalities.

The main objective of our efforts has been to highlight the importance of the social value issues inherent in the choice of any safety goal. One question that played an important role in the development of safety goals was whether current risk levels from other hazards or competing energy technologies provide meaningful benchmarks against which to set standards for nuclear power. On the basis of risk perception research, we have argued that comparisons with other risks of life or risks from competing energy sources should not be a primary factor in determining safety goals. There are many different aspects that need to be considered when evaluating a technology's risk, including perceived uncertainty regarding the probabilities and consequences of mishaps, potential for catastrophe, threat to future generations, and potential for triggering social disruption. Nuclear power is unique in many of these respects. Without an explicit logic for comparing qualitatively different risks, comparisons with other hazardous activities or technologies cannot serve as definitive guidelines for safety goals.

One question that the safety goal effort has forced us to consider in detail is whether to place special emphasis on avoiding large accidents (Slovic, Lichtenstein & Fischhoff, in press). Although psychometric studies and other surveys have pinpointed perceived catastrophic potential as a major public concern, further investigation indicates that the alpha model, the model most often proposed for incorporating risk aversion into safety goals, is incorrect. According to this model, the seriousness or social impact of losing N lives in a single accident should be modeled by the function N^{α}, where α is greater than 1.0. By attributing greater social disruption to large accidents, this model implies that small accidents may be tolerable but that extra money and effort should be expended to prevent or mitigate large accidents. Because the relationship is exponential, the spectre of low probability, catastrophic accidents can come to dominate all other considerations.

Research indicates that the alpha model is oversimplified and invalid; the societal costs of an accident cannot be modeled by any simple function of N. Rather, accidents are signals containing information about the nature and controllability of the risks involved. As a result, the perceived seriousness of an accident is often determined more by the message it conveys than by its actual toll of death and destruction. An accident will have relatively little societal impact beyond that of its direct casualties if it occurs as a result of a familiar, well understood process with little potential for recurrence or catastrophe. In contrast, an accident that causes little direct harm may have immense consequences if it increases the judged probability or seriousness of future accidents. The relationship between signal potential, accident seriousness, and the characteristics of a hazard (Figure 3) may help predict the seriousness of various mishaps.

As a case in point, the concept of accidents as signals helps explain society's strong response to some nuclear power mishaps. Because reactor risks are perceived as poorly understood and catastrophic, accidents with few direct casualties may be seen as omens of disaster, thus producing indirect or "ripple" effects resulting in immense economic costs to the industry or society. One implication of signal value is that safety goals should consider these indirect costs. A second implication is that great effort and expense might be warranted to minimize the possibility of small but frightening reactor accidents.

A final general question, which occurs with the safety goals and which may be the fundamental question motivating risk perception research is: should policy respond to public fears that experts see as unjustified? This question is currently being argued before the U.S. Supreme Court in the form of a (disputed) ruling that the undamaged reactor at Three Mile Island (there are two reactors there) cannot be restarted until the NRC has considered the effects of restart on the psychological health and well being of neighboring residents. Most experts believe that public fears of restart are groundless.

There are many reasons for laypeople and experts to disagree. These include misunderstanding, miscommunication, and misinformation (Fischhoff, Slovic & Lichtenstein, 1981; 1983). Discerning the causes underlying a particular disagreement requires careful thought, to clarify just what is being talked about and whether agreement is possible given the disputants' differing frames of reference. Also needed is careful research, to clarify just what it is that the various parties know and believe. Once the situation has been clarified, the underlying problem can be diagnosed as calling for a scientific, educational, semantic, or political solution.

Risk questions are going to be with us for a long time. For a society to deal with them wisely, it must understand their subtleties. We believe that research within the framework of behavioral decision theory is essential to achieving this understanding.

References

Brown, R. A. & Green, C. H. Precepts of safety assessments. Journal of the Operational Research Society, 1980, 11, 563-571.

Douglas, M. & Wildavsky, A. Risk and culture. Berkeley and Los Angeles: University of California Press, 1982.

Earle, T. C. & Lindell, M. K. Public perception of industrial risks: A free-response approach. Proceedings of the 1982 meeting of the Society for Risk Analysis, in press.

Fischhoff, B. Cognitive and institutional barriers to "informed consent". In M. Gibson (Ed.), Risk, consent, and air. Totowa, New Jersey: Rowman & Allanheld, in press-a.

Fischhoff, B. Informed consent for transient nuclear workers. In R. Kasperson and R. W. Kates (Eds.), Equity issues in nuclear waste disposal. Cambridge, Mass.: Oelgeschlager, Gunn & Hain, in press-b.

Fischhoff, B. A theory of safety standards. Management Science, in press-c.

Fischhoff, B. Safety goals for nuclear power. Journal of Policy Analysis and Management, 1983, 2, 559-575.

Fischhoff, B. & MacGregor, D. Judged lethality. Risk Analysis, in press.

Fischhoff, B., Slovic, P. & Lichtenstein, S. Lay foibles and expert fables in judgments about risk. In T. O'Riordan and R. K. Turner (Eds.), Progress in resource management and environmental planning, Vol. 3. Chichester: Wiley, 1981.

Fischhoff, B., Slovic, P. & Lichtenstein, S. The "public" vs. the "experts": Perceived vs. actual disagreement about the risks of nuclear power. In V. Covello, G. Flamm, J. Rodericks, and R. Tardiff (Eds.), Analysis of actual vs. perceived risks. New York: Plenum, 1983.

Fischhoff, B., Watson, S. & Hope, C. Defining risk. Decision Research Report 82-15, 1983.

Gardner, G. T., Tiemann, A. R., Gould, L. C., DeLuca, D. R., Doob, L. W. & Stolwijk, J. A. J. Risk and benefit perceptions, acceptability judgments, and self-reported actions toward nuclear power. Journal of Social Psychology, 1982, 116, 179-197.

Gibson, M. (Ed.), Risk, consent, and air. Totowa, New Jersey: Rowman & Allanheld, in press.

Green, C. H. Risk: Attitudes and beliefs. In D. V. Canter (Ed.), Behaviour in fires. Chichester: Wiley, 1980.

Green, C H. & Brown, R. A. Through a glass darkly: Perceiving perceived risks to health and safety. Research paper, School of Architecture, Duncan of Jordanstone College of Art/University of Dundee, Scotland, 1980.

Johnson, E. J. & Tversky, A. Representations of perceptions of risks. Journal of Experimental Psychology: General, in press.

Kahneman, D. & Tversky, A. Prospect theory: An analysis of decisions under risk. Econometrica, 1979, 47, 262-291.

Lindell, M. K. & Earle, T. C. How close is close enough: Public perceptions of the risks of industrial facilities. Unpublished manuscript, Battelle Human Affairs Research Center, Seattle, 1982.

Macgill, S. M. Exploring the similarities of different risks. Working paper, The School of Geography, University of Leeds, 1982.

McNeil, B. J., Pauker, S. G., Sox, H. C. & Tversky, A. On the elicitation of preferences for alternative therapies. New England Journal of Medicine, 1982, 306, 1259-1262.

Nisbett, R. E. & Ross, L. Human inference: Strategies and shortcomings of social judgment. Englewood Cliffs, N. J.: Prentice-Hall, 1980.

Otway, H. J. & Thomas, K. Reflections on risk perception and policy. Risk Analysis, 1982, 2, 69-82.

Otway, H. J. & von Winterfeldt, D. Beyond acceptable risk: On the social acceptability of technologies. Policy Sciences, 1982, 14, 247-256.

Perusse, M. Dimensions of perception and recognition of danger. Ph.D. dissertation, University of Aston, Birmingham, 1980.

Renn, O. Man, technology, and risk: A study on intuitive risk assessment and attitudes towards nuclear power. Report Jul-Spez 115, Julich, Federal Republic of Germany: Nuclear Research Center, June 1981.

Sales, J. B. The duty to warn and instruct for safe use in strict tort liability. St. Mary's Law Journal, 1982, 13, 521-586.

Slovic, P., Fischhoff, B. & Lichtenstein, S. Informing people about risk. In L. Morris, M. Mazis and I. Barofsky (Eds.), Product labeling and health risks. Banbury Report 6. Cold Spring Harbor, New York: Cold Spring Laboratory, 1980a.

Slovic, P., Fischhoff, B. & Lichtenstein, S. Facts and fears: Understanding perceived risk. In R. Schwing and W. A. Albers, Jr. (Eds.), Societal risk assessment: How safe is safe enough? New York: Plenum, 1980b.

Slovic, P., Fischhoff, B. & Lichtenstein, S. Informing the public about the risks of ionizing radiation. Health Physics, 1981a, 41, 589-598.

Slovic, P., Fischhoff, B. & Lichtenstein, S. Perception and acceptability of risks from energy systems. In A. Baum and J. E. Singer (Eds.), Advances in Environmental Psychology, Vol. 3. Hillsdale, N. J.: Erlbaum, 1981b.

Slovic, P., Fischhoff, B. & Lichtenstein, S. Response mode, framing, and information-processing effects in risk assessment. In R. Hogarth (Ed.), New directions for methodology of social and behavioral science: Question framing and response consistency. San Francisco, Calif.: Jossey-Bass, 1982a.

Slovic, P., Fischhoff, B. & Lichtenstein, S. Why study risk perception? Risk Analysis, 1982b, 2, 83-93.

Slovic, P., Fischhoff, B. & Lichtenstein, S. Characterizing perceived risk. In R. W. Kates, C. Hohenemser, and J. X. Kasperson (Eds.), Perilous progress: Technology as hazard. Cambridge, Mass.: Oelgeschlager, Gunn & Hain, in press.

Slovic, P., Lichtenstein, S. & Fischhoff, B. Modeling the societal impact of fatal accidents. Management Science, in press.

Solomon, K. A., Nelson, P. F. & Salem, S. L. Utilization of risk analysis and risk criteria: A progress report. Working draft 1194-ORNL. Santa Monica, Calif.: RAND Corp., August 1981.

Tversky, A. & Kahneman, D. The framing of decisions and the psychology of choice. Science, 1981, 211, 1453-1458.

Vlek, C. A. J. & Stallen, P. J. M. Judging risks and benefits in the small and in the large. Organizational Behavior and Human Performance, 1981, 28, 235-271.

von Winterfeldt, D. & Edwards, W. Patterns of conflict about risky technologies. Unpublished manuscript, 1983.

von Winterfeldt, D., John, R. S. & Borcherding, K. Cognitive components of risk ratings. Risk Analysis, 1981, 1, 277-287.

Figure Captions

1. Hazard locations on Factors 1 and 2 derived from the interrelationships among 16 risk characteristics. Each factor is made up of a combination of characteristics, as indicated by the lower diagram. Source: Slovic, Fischhoff and Lichtenstein (in press).

2. Attitudes towards regulation of the hazards in Figure 1. The larger the point, the greater the desire for strict regulation to reduce risk.

3. Relation between signal potential and risk characterization for 30 hazards in Figure 1. The larger the point, the greater the degree to which an accident involving that hazard was judged to "serve as a warning signal for society, providing new information about the probability that similar or even more destructive mishaps might occur within this type of activity." Source: Slovic, Lichtenstein and Fischhoff (in press).

4. Tree representation of causes of death. Source: Johnson and Tversky (in press).

Table 1

Lethality Judgments with Different Response Modes (Geometric Means)

	Death rate per 100,000 afflicted				
Malady	Estimated lethality rate	Estimated number who die	Estimated survival rate	Estimated number who survive	Actual lethality rate
Influenza	393	6	26	511	1
Mumps	44	114	19	4	12
Asthma	155	12	14	599	33
Veneral disease	91	63	8	111	50
High blood pressure	535	89	17	538	76
Bronchitis	162	19	43	2111	85
Pregnancy	67	24	13	787	250
Diabetes	487	101	52	5666	800
Tuberculosis	852	1783	188	8520	1535
Automobile accidents	6195	3272	31	6813	2500
Strokes	11011	4648	181	24758	11765
Heart attacks	13011	3666	131	27477	16250
Cancer	10889	10475	160	21749	37500

The four experimental groups were given the following instructions:
(a) Estimate lethality rate: for each 100,00 people afflicted, how many die?
(b) Estimate number who die: X people were afflicted, how many died?
(c) Estimate survival rate: for each person who died, how many were afflicted but survived?
(d) Estimate number who survive: Y people died, how many were afflicted but did not die?
Responses to questions (b), (c), and (d) were converted to deaths per 100,000 to facilitate comparisons.

Source: Fischhoff & MacGregor, in press.

Table 2
Constructs Elicited by Means of the Repertory Grid Technique

ORIGIN OF DANGER
- Natural/man-made
- Human cause/no human cause
- Blame assignable/no blame assignable
- Self responsible/self not responsible
- internal/external

CHARACTERISTICS OF HAZARDS
- Necessary/unnecessary activity
- Occupational/not occupational
- Potential/present
- Near/far
- Moving/stationary
- Slow/fast event
- Specific/non-specific location
- Open/enclosed
- Large/small concentration of people

THREAT
- Frequent/infrequent occurrence
- High/low risk of accident
- Most dangerous/least dangerous
- Safe/unsafe
- Sudden/continuous threat

CONSEQUENCES
- Major/minor
- Large/small consequence
- Fatal/survivable
- Many killed/few killed
- Many affected/few affected
- Personal/impersonal
- Instantaneous/long-term consequence
- Reversible/irreversible
- Painful/painless

HUMAN INTERVENTION
- Own control/out of control
- Rely on others/rely on self
- Avoidable/unavoidable
- Preventable/unpreventable
- Precautions/no precautions
- Foreseeable/unforeseeable
- Easy/difficult to escape

REACTIONS
- Aware/unaware of danger
- Sleeping/awake
- Familiar/unfamiliar
- Ugly-hideous/not ugly
- Scaring/not scaring
- Worry-concern/non-worry, unconcern
- Acceptance/non-acceptance
- Panic-chaos/orderly-calm
- Public reaction/no public reaction

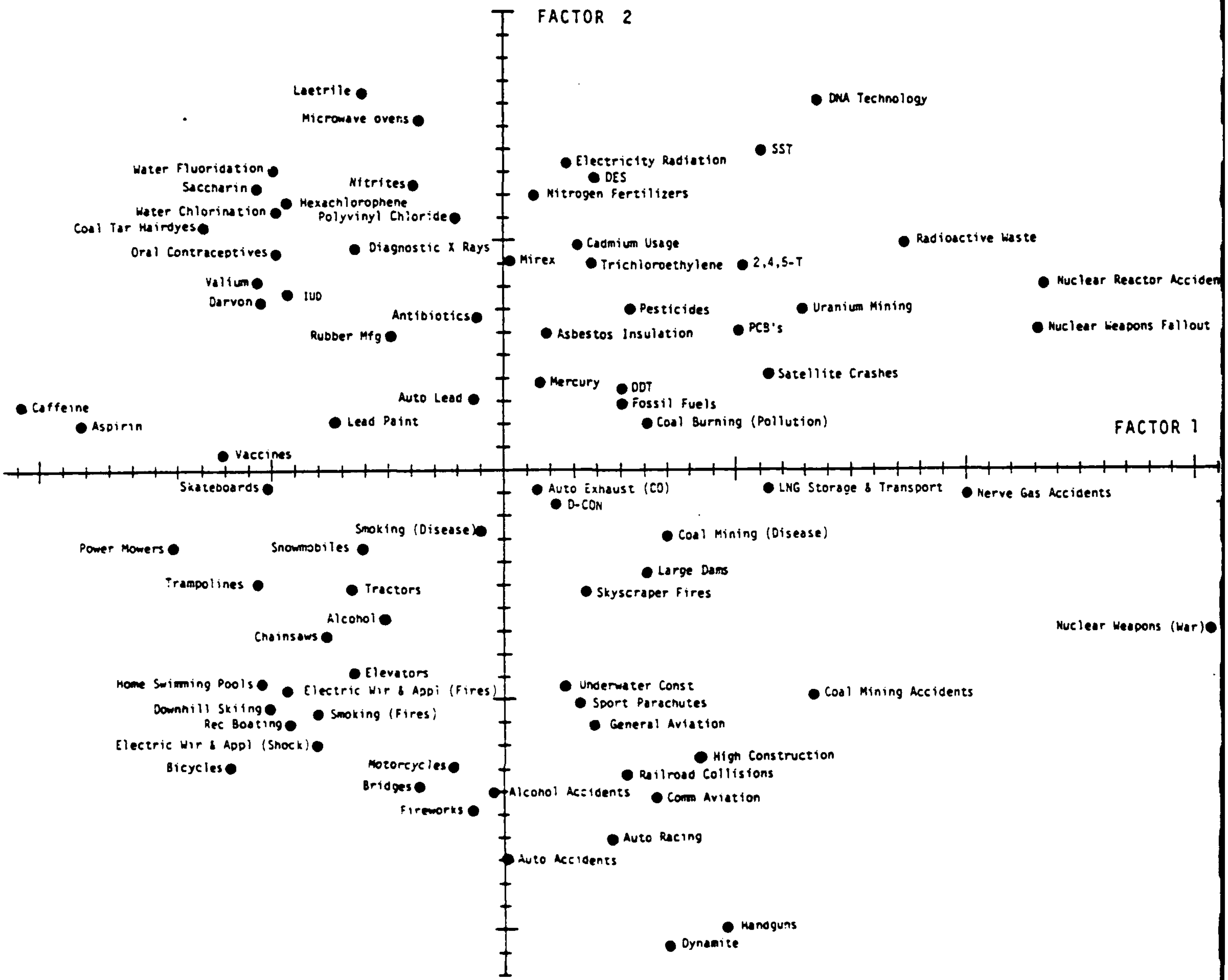
Source: Perusse, 1980

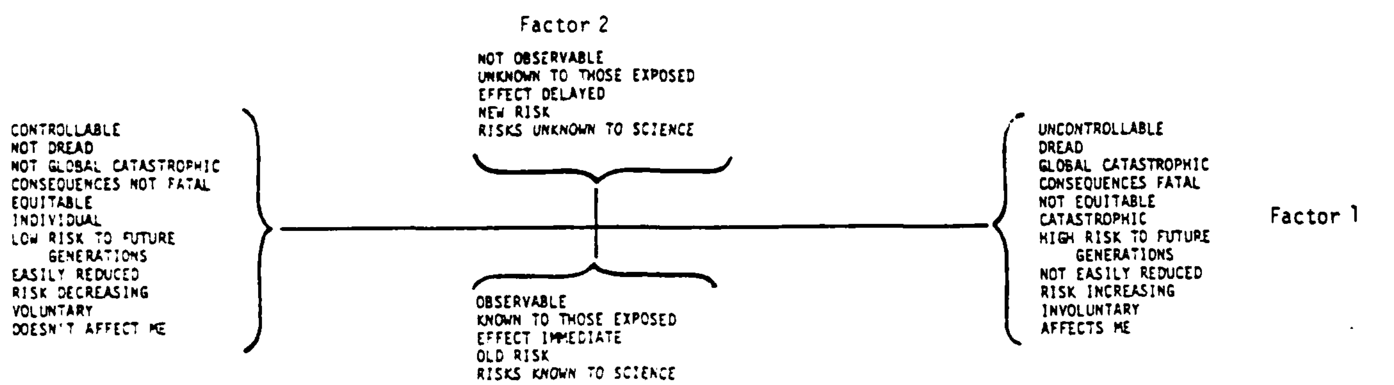
Table 3

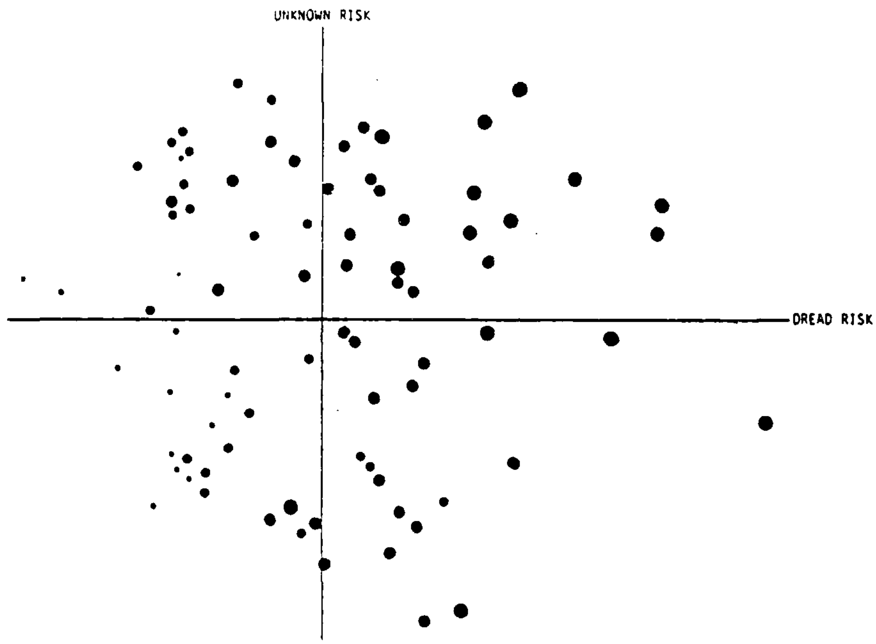
Conditions Justifying the Development of Safety Standards

1. When predictability is important.
2. When one need not choose a single best option.
3. When a single (standardizable) feature captures the most important aspect of a category.
4. When the standard accurately postdicts past decisions and predicts future ones.
5. When one wants to make a statement to reflect the goals of policy makers (who assume the symbolic standard will be reasonably compromised by those who apply it).
6. When one hopes to shape the set of future options.
7. When the decision process leading to the standard is of higher quality than could be maintained in numerous specific decisions.

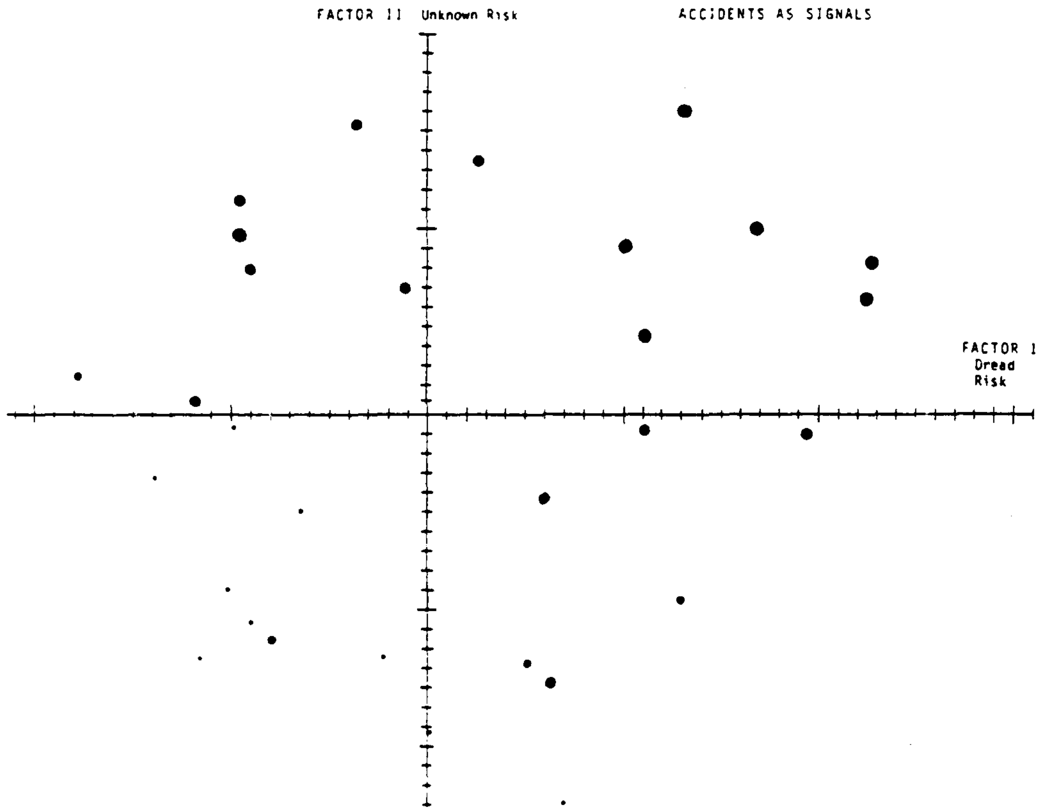
Source: Adapted from Fischhoff (in press-c).

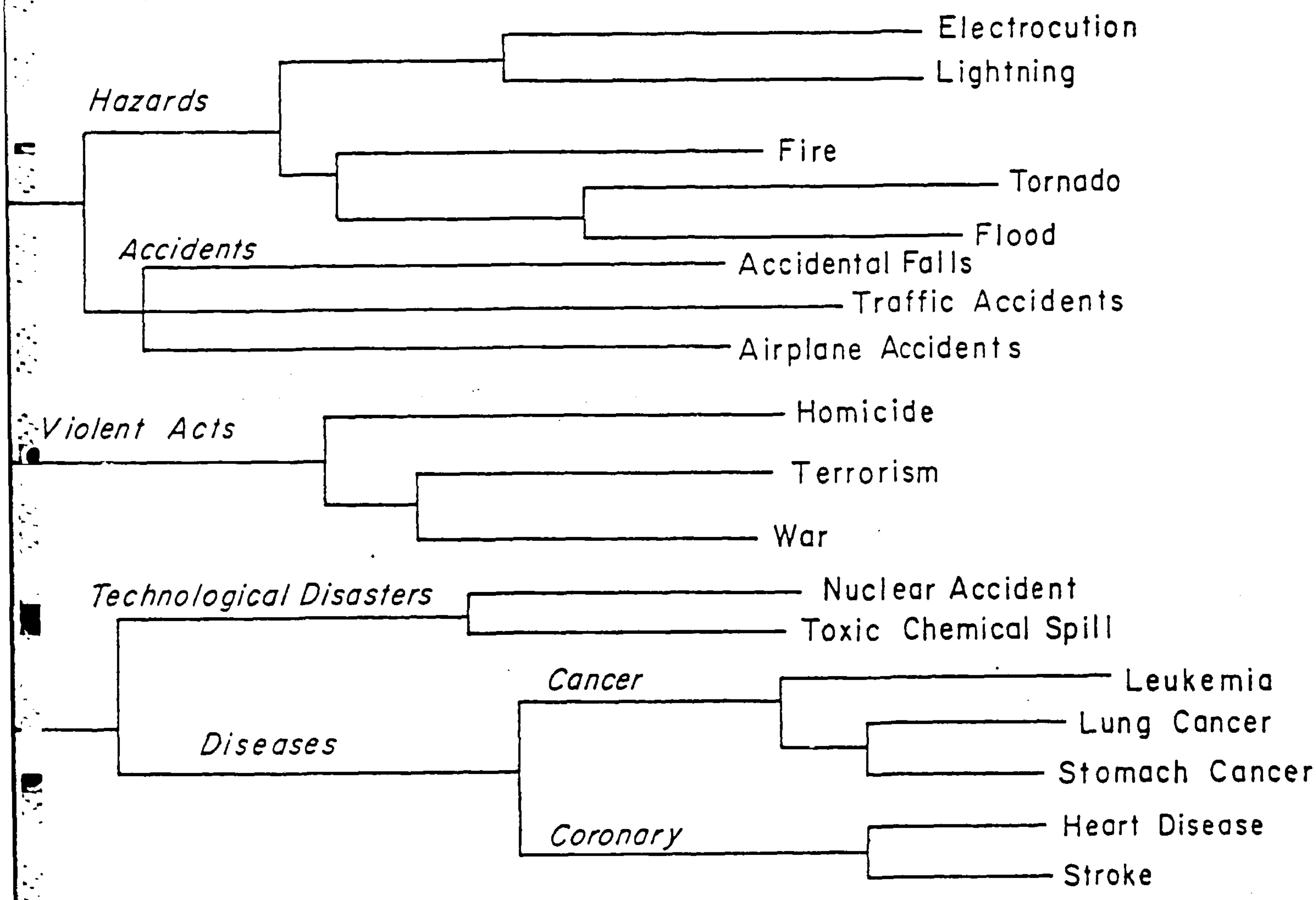

FACTOR 2
FACTOR 1
Laetrile
Microwave ovens
DNA Technology
SST
Electricity Radiation
DES
Nitrogen Fertilizers
Water Fluoridation
Saccharin
Nitrites
Hexachlorophene
Water Chlorination
Polyvinyl Chloride
Coal Tar Hairdyes
Oral Contraceptives
Diagnostic X Rays
Mirex
Cadmium Usage
Trichloroethylene
2,4,5-T
Radioactive Waste
Nuclear Reactor Acciden
Valium
IUD
Darvon
Antibiotics
Pesticides
Uranium Mining
Rubber Mfg
Asbestos Insulation
PCB's
Nuclear Weapons Fallout
Satellite Crashes
Mercury
DDT
Fossil Fuels
Auto Lead
Caffeine
Aspirin
Lead Paint
Coal Burning (Pollution)
Vaccines
Skateboards
Auto Exhaust (CO)
D-CON
LNG Storage & Transport
Nerve Gas Accidents
Smoking (Disease)
Power Mowers
Snowmobiles
Coal Mining (Disease)
Large Dams
Trampolines
Tractors
Skyscraper Fires
Alcohol
Chainsaws
Nuclear Weapons (War)
Elevators
Home Swimming Pools
Electric Wir & Appl (Fires)
Underwater Const
Sport Parachutes
Coal Mining Accidents
Downhill Skiing
Rec Boating
Smoking (Fires)
General Aviation
Electric Wir & Appl (Shock)
High Construction
Bicycles
Motorcycles
Railroad Collisions
Bridges
Alcohol Accidents
Comm Aviation
Fireworks
Auto Racing
Auto Accidents
Handguns
Dynamite


Factor 2
NOT OBSERVABLE
UNKNOWN TO THOSE EXPOSED
EFFECT DELAYED
NEW RISK
RISKS UNKNOWN TO SCIENCE
CONTROLLABLE
NOT DREAD
NOT GLOBAL CATASTROPHIC
CONSEQUENCES NOT FATAL
EQUITABLE
INDIVIDUAL
LOW RISK TO FUTURE GENERATIONS
EASILY REDUCED
RISK DECREASING
VOLUNTARY
DOESN'T AFFECT ME
UNCONTROLLABLE
DREAD
GLOBAL CATASTROPHIC
CONSEQUENCES FATAL
NOT EQUITABLE
CATASTROPHIC
HIGH RISK TO FUTURE GENERATIONS
NOT EASILY REDUCED
RISK INCREASING
INVOLUNTARY
AFFECTS ME
Factor 1
OBSERVABLE
KNOWN TO THOSE EXPOSED
EFFECT IMMEDIATE
OLD RISK
RISKS KNOWN TO SCIENCE



UNKNOWN RISK
DREAD RISK


FACTOR II Unknown Risk
ACCIDENTS AS SIGNALS
FACTOR I
Dread
Risk



Hazards
Accidents
Electrocution
Lightning
Fire
Tornado
Flood
Accidental Falls
Traffic Accidents
Airplane Accidents
Violent Acts
Homicide
Terrorism
War
Technological Disasters
Nuclear Accident
Toxic Chemical Spill
Diseases
Cancer
Leukemia
Lung Cancer
Stomach Cancer
Coronary
Heart Disease
Stroke

Final Report

Contract N00014-82-C-0643, Project No. NR 274-347

(FRAMING AND EVALUATION OF RISK ANALYSIS)

Principal Investigator
Paul Slovic

Decision Research
A Branch of Perceptronics
1201 Oak Street
Eugene, Oregon 97401
(503) 485-2400

March 30, 1984

Final Report:

Our research to date under this contract has been organized around three separate projects. These are:

(a) Studies of response mode and framing effects

(b) Studies of the acceptability of decision making methods

(c) Preparation of the chapter on Decision Making for the revised Handbook of Experimental Psychology

A brief review of the status of work on these projects follows.

1.0 Response Mode and Framing Effects

1.1 Preference Reversals

Preference reversals illustrate, in a dramatic way, the strong influence of information-processing factors on perception and evaluation of risky options. This direction of research originated with studies by Slovic and Lichtenstein in 1968 and 1971 showing that evaluation of gambles depended greatly on response mode. Subjects were presented with pairs of gambles, one featuring a high probability of winning a modest amount of money (the P bet) and one featuring a low probability of winning a large amount (the $ bet). The typical finding was that people often chose the P bet but assigned a larger monetary value (buying price, selling price) to the $ bet. This behavior is of interest because it violates almost all theories of preference, including expected utility theory. Lichtenstein and Slovic have explained such reversals by proposing that the different response modes trigger different mental operations for processing the information in a gamble.

Economists have been resistent to the notion of preference reversals, perhaps because of the disturbing implications for economic theory. They have conducted several major studies attempting to show that, under proper experimental conditions (e.g., proper motivation and instructions) the reversals would disappear. The reversal effect has survived in each of these studies, yet new studies keep being designed in hopes of disproving the phenomenon.

Following the publication of two such studies in the June, 1982 issue of the American Economic Review, Slovic and Lichtenstein decided to write a rejoinder, pointing out that preference reversals had been obtained under a wide variety of rigorous conditions and it was now time for economists to put effort into considering their practical and theoretical implications. Some economists, such as Thaler and Arrow have already begun to do this, as our paper indicated. The Slovic and Lichtenstein paper was published in the American Economic Review. A copy is attached to this progress report as Appendix A.

In addition, Amos Tversky and Paul Slovic have been collaborating on some new studies of preference reversals that are producing some striking results, broadening the scope and implications of response-mode effects. Tversky and Slovic are using the simplest form of gamble, probability P to win $X. Subjects are given pairs of gambles, such as:

A (P bet)	vs.	B ($ bet)
35/36 to win $4.00		1/36 to win $16.00

Tversky and Slovic have obtained the usual form of reversal when they ask for prices (bids) and choices. The price attached to the $ bet exceeds the price of the P bet 85-90% of the time, but the P bet is chosen about 60-70% of the time. A new twist is that they also asked people to rate the attractiveness of playing each bet on a 1-20 scale. This rating produced an overwhelming dominance of P bets over their corresponding $ bets (the P bet had the higher attractigveness rating almost 90% of the time). So the results, in terms of percent superiority of the P bet over the $ bet within a pair are as follows:

Pricing	Choice	Attractiveness Rating
10-15%	60-70%	80-90%

Reversals were found not only between pricing and choice, but between ratings and prices and even between choice and ratings. They have explained this pattern of results in terms of the notion of stimulus-response compatibility (or codeability) and have begun to develop a theory of this process. For example, probabilities have a precise translation into attractiveness: high probabilities of winning are highly attractive; low probabilities of winning are not attractive. Payoffs have a less clear relation to attractiveness. Hence probabilities are weighted far more heavily than payoffs when judging attractiveness. Similar processes, in the opposite direction, cause payoffs to dominate probabilities when bets are evaluated in monetary terms. These results broaden our understanding of response mode effects and their implications for theories of preference and utility. A draft report of this study has been completed and is attached to this progress report as Appendix B.

1.2 Judged Lethality

Baruch Fischhoff and Donald MacGregor have completed another study of response mode effects, pertaining to the perception of lethality from various causes of death. Four formally equivalent modalities were used to elicit laypeople's beliefs regarding the lethality of various potential causes of death. Results showed that respondents had an articulated core of beliefs about lethality which yielded similar orderings of maladies by lethality regardless of the elicitation modality used. Moreover, this subjective ordering was fairly similar to that revealed by public health statistics. However, the absolute estimates of lethality produced by the different modalities varied enormously. Depending upon the modality used, respondents were seen to greatly overestimate or greatly underestimate lethality. The results appear to have important implications for the elicitation and communication of risk information. A complete report of this study is appended to this proposal (Appendix C). We should note that the original data in this study were collected under an earlier contract (from ARPA). Work under the current proposal has involved a complete reanalysis of the data and rewriting of the paper from the perspective of response mode and presentation effects.

1.3 Studies of Framing

We have completed an extensive series of studies that test some of the basic principles of Kahneman and Tversky's Prospect Theory, particularly the editing and framing components of their theory. The basic stimulus used in these studies is the "civil defense problem" shown in Figure 1. As the figure indicates, the same decision problem can be viewed from three different perspectives. According to Prospect Theory, perspectives I and II should lead to a preference for the gamble and Frame III should lead to a preference for the sure loss.

We have, to date, completed eight sub-studies , in which we have varied the parameters, context, wording, instructions and order of presentation of the frames. We have found that people tend to adopt Frames I and II as the dominant perspectives, leading them to select the gamble over the sure loss. Attempts to induce people to adopt Frame III appear not to have been successful. In other words, people do not appear to be able to absorb the 50 lives lost into a neutral (or status quo) reference point. Another result is that people's introspective judgments as to the relative naturalness of each frame do not seem to be related to their preferences, contrary to predictions from Prospect Theory.

We believe that we are tracking something very important in these studies. Although Kahneman and Tversky have clearly shown the importance of the way that a decision problem is framed, they always imposed the frame on the subjects in their studies. Little is known regarding the frames people adopt when they are free to view a

Figure 1

Decision Framing: Three perspectives on a civil defense problem

A civil defense committee in a large metropolitan area met recently to discuss contingency plans in the event of various emergencies. One emergency threat under discussion posed two options, both involving some loss of life.

Option A: Carries with it a .5 probability of containing the threat with a loss of 40 lives and a .5 probability of losing 60 lives. It is like taking the gamble:

.5 lose 40 lives

.5 lose 60 lives

Option B: Would result in the loss of 50 lives:

lose 50 lives

These options can be presented under three different frames:

I. This is a choice between a 50-50 gamble (lose 40 or lose 60 lives) and a sure thing (the loss of 50 lives).

II. Whatever is done at least 40 lives will be lost. This is a choice between a gamble with a 50-50 chance of either losing no additional lives or losing 20 additional lives (A) and the sure loss of 10 additional lives (B).

III. Option B produces a loss of 50 lives. Taking Option A would mean accepting a gamble with a .5 chance to save 10 lives and a .5 chance to lose 10 additional lives.

decision problem from multiple perspectives. Our investigations thus far suggest that frames may sometimes be surprisingly hard to manipulate. Furthermore, there is a disturbing lack of correspondence between the frames they do adopt, their subsequent choices, and the predictions that Prospect Theory makes, given these frames. A better understanding of the framing process is very much needed. A report on our recent studies in this area is being prepared.

2.0 Studies of the Acceptability of Risk-Analysis Methods

We have conducted several studies related to the perceived acceptability of risk analysis as a decision making method. The first of these studies was a large, between-subjects multi-factor design that took an initial look at the influence of a number of factors on people's judgments of several forms of risk analysis, including cost-benefit analysis and expected-value risk analysis. The latter method differs from cost-benefit analysis only in that it does not explicitly trade off values, but instead calls for a deliberative choice on the part of a decision maker. Under the current contract, we have been following up some key methodological and substantive findings from that study.

A key finding from this research is that people prefer expected value risk analysis over cost-benefit analysis. One suggestion from these results is that, despite the claim of cost-benefit analysis that it makes value tradeoffs expicit, people may prefer to have decision makers intuitively and holistically arrive at a choice rather than abide solely by the outcome of an explicit, quantitative analysis. In other words, an approach that uses risk analysis as an input into an intuitive deliberative decision-making process will be viewed more favorably by the lay public than a purely intuitive or purely analytic process.

Another of our studies has looked closely at the role of risk analysis in guiding a decision maker regarding the choice of whether or not to expose people to a hazardous consumer procudt. The decision maker had access to a consumer poll as well as to the risk analysis. There were four cells in the design of the study, resulting from a combination of two factors: the outcome of the analysis could either favor or not favor using the product and the outcome of the poll could either favor or not favor the product. In each of the four conditions, subjects were asked to judge how much weight they would give to the analysis and to the poll in arriving at a decision. Preliminary results suggest that, when the poll opposed the action, the analysis was acceptable only if it corroborated the preferences in the poll. However, when the poll favored the risky action, the balance of costs and benefits was judged an aceptable basis for making the decision.

In other studies, we have found that cost/benefit analysis is judged more appropriate to deal with decisions involving economic matters than decisions in which people's lives and health are at stake. We have also found that people are able to separate the monetizable aspects of a risk decision from the non-monetizable ones and they seem to want the non-monetizable ones included in the analysis. This suggests that an approach, based on multi-attribute utility theory, that explicitly considers and blends monetizable and non-monetizable outcomes, may be judged rather favorably.

In sum, our studies to date represent a first step towards developing an understanding of the ways that people respond to the interplay between analytic/mechanistic and intuitive/deliberative elements in decision making approaches. Our method for studying this topic seems tractable and produces results that seem to make sense in terms of the kinds of debates and controversies that are currently going on in society's attempts to manage risks.

3.0 Chapter on Decision Making

The Handbook of Experimental Psychology was originally published in 1951, edited by S. S. Stevens of Harvard University. This landmark volume contained 36 chapters on all major aspects of experimental psychology. Decision Making was not included as a chapter because its empirical study was in its infancy at that time. During the past three decades much has happened to change the face of experimental psychology and the field of decision making. The original handbook is badly out of date. The field of decision making has burgeoned into a major theoretical and empirical line of inquiry. Accordingly, a revised edition of the Handbook is being prepared, under the editorship of Richard Atkinson, Gardner Lindzey, Duncan Luce,and Richard Herrnstein. We were asked to write the chapter on Decision Making for this revised Handbook. We have completed the chapter and submitted a copy to our project monitor. The table of contents for this chapter is given on pages 11 and 12.

4.0 Other Work

We were invited to prepare a paper for the ninth conference on subjective probability, utility, and decision making held in The Netherlands in August, 1983. A copy of that paper, "Behavioral Decision Theory Perspectives on Risk and Safety," will be published in Acta Psychologica and is attached as Appendix D.

Appendices to the Final Report for

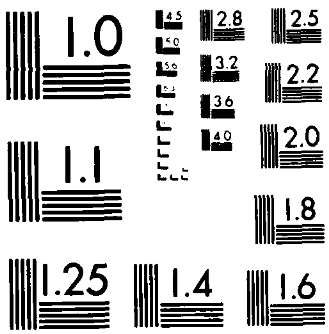
Contract N00014-82-C-0643

(FRAMING AND EVALUATION OF RISK ANALYSIS)

AD-A148 399 COMPATIBILITY EFFECTS AND PREFERENCE REVERSALS(U) PERCEPTRONICS INC WOODLAND HILLS CA A TVERSKY ET AL. 21 AUG 84 PFTR-1127-84-8 N00014-82-C-0643 2/2

UNCLASSIFIED F/G 5/10 NL

END FILMED DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Preference Reversals: A Broader Perspective

By PAUL SLOVIC AND SARAH LICHTENSTEIN*

Two papers recently published in this *Review*, the first by Werner Pommerehne, Freidrich Schneider, and Peter Zweifel (1982) and the second by Robert Reilly (1982), reexamined the preference reversal phenomenon. Preference reversals occur when individuals are presented with two gambles, one featuring a high probability of winning a modest sum of money (the *P* bet), the other featuring a low probability of winning a large amount of money (the $ bet). The typical finding is that people often choose the *P* bet but assign a larger monetary value to the $ bet. This behavior is of interest because it violates almost all theories of preference, including expected utility theory.

The studies by Pommerehne et al. and Reilly were based on an earlier paper appearing in this *Review* by David Grether and Charles Plott (1979). All three of these investigations have followed the same general design, motivated by a healthy skepticism of the phenomenon and a belief that, examined under proper conditions, it might disappear. Thus Grether and Plott took great pains to correct what they saw as deficiencies in the original psychological experiments by ourselves (1971, 1973) and Harold Lindman (1971). Specifically, Grether and Plott used two monetary incentive systems to heighten motivation, substituted a different probability device for deciding the outcomes of the bets, controlled for income and order effects, and tested for indifference and the influence of strategic or bargaining effects. To their surprise, preference reversals remained much in evidence, despite their careful attempts to create conditions that would minimize or eliminate them.

Pommerehne et al., not satisfied with the stringency of Grether and Plott's controls, attempted to increase motivation by raising the face value of the payoffs and creating differences in expected value between the *P* and $ bets in a pair. They, too, found a substantial proportion of reversals, leading them to conclude: "Even when the subjects are exposed to strong incentives for making motivated, rational decisions, the phenomenon of preference reversal does not vanish" (p. 573).

Reilly was also skeptical of the adequacy of Grether and Plott's controls. To maximize subjects' understanding of the task, he conducted his study within small groups where questions could readily be asked of the experimenter. The money at risk was placed on a desk in front of the subject and the size of potential losses in the gambles was increased to enhance motivation. Finally, some subjects were shown the expected values for all gambles and were given a description of the expected-value concept. Although the rate of preference reversals was somewhat lower than that observed by Grether and Plott, the phenomenon persisted to a substantial extent. Reilly conceded that these results provided "further confirmation of preference reversal as a persistent behavioral phenomenon in situations where economic theory is generally applied" (p. 582). Nevertheless, he maintained the hope that further strengthening of monetary incentives and provision of additional information to the subjects would make this troublesome phenomenon disappear, thus salvaging preference theory:

> Should sufficiently large reductions be achievable, we might consider adopting the premise that individuals are likely to be consistent in making decisions that *matter* to them when the principle characteristics of the alternatives are sufficiently comprehended. Applied to such cases, standard preference theory would then require little modification. [p. 582]

*Decision Research, A Branch of Perceptronics, 1201 Oak Street, Eugene, Oregon 97401. The work was supported by the Office of Naval Research under Contract N00014-82-C-0643 to Perceptronics, Inc. We thank Don MacGregor, Amos Tversky, and two anonymous reviewers for comments on an earlier draft.

As researchers who have studied preference reversals and related problems of rational choice for quite some time, we have several concerns about the direction this research seems to be taking. Certainly a phenomenon such as preference reversals should be subjected to rigorous tests such as those administered by Grether and Plott, Pommerehne et al., and Reilly. These studies have been valuable in demonstrating the robustness of the effect. However, there is a substantial body of research on preference reversals within the psychological literature that is being neglected here. Moreover, reversals can be seen not as an isolated phenomenon, but as one of a broad class of findings that demonstrate violations of preference models due to the strong dependence of choice and preference upon information processing considerations. In this paper we shall describe relevant psychological work in order to broaden the perspective on preference reversals.

I. History

Readers of the papers by Pommerehne et al. and Reilly would hardly know there was considerable scrutiny of preference reversals prior to the publication by Grether and Plott. In fact, a number of studies preceded Grether and Plott, most of which employed multiple experiments and conditions designed to test the robustness of the effect. Additional studies have appeared subsequently. Each of these studies has observed substantial frequencies of reversals.

The first study designed to elicit reversals was our 1971 article. The impetus for this study was our observation in our earlier 1968 article that choices among pairs of gambles appeared to be influenced primarily by probabilities of winning and losing, whereas buying and selling prices were primarily determined by the dollar amounts that could be won or lost. When subjects found a bet attractive, their prices correlated predominantly with the amount to win; when they disliked a bet, their prices correlated primarily with the amount that could be lost. This pattern of correlations was explained as the result of a starting point (anchoring) and adjustment procedure used when setting prices. Subjects setting a price on an attractive gamble appeared to start with the amount to win and adjust it downward to take into account the probability of winning and losing, and the amount that could be lost. The adjustment process was relatively imprecise, leaving the price response greatly influenced by the starting point payoff. Choices, on the other hand, appeared to be governed by different rules.

In our 1971 article, we argued that, if the information in a gamble is processed differently when making choices and setting prices, it should be possible to construct pairs of gambles such that people would choose one member of the pair but set a higher price on the other. We proceeded to construct a small set of pairs that clearly demonstrated this predicted effect.[1] Following this, a second study was conducted to examine the strength of the reversal effect as a function of the characteristics of the bet pairs. Forty-nine pairs of bets were constructed, all constrained by the requirement that the *P* bet had a high probability of winning a modest amount and the $ bet had a low to moderate probability of winning a large amount. Despite these constraints, the pairs differed significantly in the degree to which they elicited predictable reversals. The ideal bet pair for observing reversals had a larger $ bet loss than a *P* bet loss (facilitating choice of the *P* bet) and a large $ bet win relative to the *P* bet win (facilitating a higher price for the $ bet). For example, the bet with the most predicted reversals was: *P* bet, 9/12 to win $1.20 and 3/12 to lose $.10; $ bet, 3/12 to win $9.20 and 9/12 to lose $2.00. We concluded this initial study by noting that reversals were of interest not only because they violated theories of rational choice, but because of the insight they revealed about the nature of judgment and decision processes.

Our 1968 article also noted that the close dependence of pricing responses on a gamble's payoffs could explain a finding that had

[1] Contrary to the explanation by Reilly, the act of choosing the *P* bet but setting a higher price on the $ bet is not called a *predicted* reversal simply because "In all experiments reversal of *P* bets has been more frequent than reversal of $ bets..." (Reilly, 1982, p. 577, fn. 2).

puzzled Lindman (1965) in his doctoral dissertation. Lindman's subjects gave selling prices for gambles and also made paired-comparison choices among triplets of these gambles. He noted that the prices were ordered almost perfectly according to the payoffs, whereas the orderings derived from choices were not. Lindman (1971) subsequently performed five studies designed to determine whether this sort of inconsistency would be influenced by the number of gambles within the choice set, the possibility of comparing gambles directly when deciding upon selling prices, variations in the way that probabilities were displayed, and variations in the amount of prior practice or experience. Although the experience factor had some effect, the general results across conditions were in close agreement with our own.

Problems of motivation and understandability were of concern right from the beginning of these studies. Experiment III of our original paper (1971) allowed college student subjects to win up to $8, a significant amount for an hour's work in 1969. Each subject was run individually, with lengthy and careful instructions. Prices and choices were obtained three times for each pair of bets. The third time, subjects were reminded what their earlier answers had been and were asked to make a careful, final response. The bets were actually played and subjects were paid as a function of their winnings. Results for these carefully trained and financially motivated subjects showed a substantial proportion of predicted reversals. Recognizing the importance of motivation and the need to test nonstudent subjects, we went to considerable effort to replicate the initial studies on the floor of a casino in downtown Las Vegas.[2] There the players could set the value of their chips at $.05, $.10, $.25, $1, or $5. No players ever chose $1 or $5, but even for the $.10 chips, a typical $ bet offered either a win or a loss of $8 on a single play. One new feature of the design was the addition of gambles having negative expected values. The experiment attracted 44 players, many of whom were highly educated professionals. Reversals of preference were frequent and widespread across players, even for the negative expected value gambles, for which strategic tendencies to overprice the bets would have worked against the reversal phenomenon.

Robert Hamm (1979) was another researcher who tried hard to make the reversal phenomena disappear—and did not. His extensive study examined the stability of reversals over time, in the face of experience, practice, forced introspection or discussion, and advice to adopt an intuitive or analytic approach to the task. The order of stimulus sets and tasks was carefully counterbalanced. Hamm found that the reversal effect was replicated under all these conditions. Task order had no effect, nor did emphasis on analytic or intuitive processes. Discussion about one's decision strategies actually increased the tendency towards reversals, countering the hypothesis that if people were given greater opportunity to think about their strategies, the preference reversal phenomenon would disappear.

John Mowen and James Gentry (1980) studied preference reversals in a quite different context—that of new product development. Their subjects were undergraduate students of marketing and consumer behavior. They also extended previous research by comparing individual vs. group decisions. The stimuli were hypothetical products, defined according to probability of success and failure, and the projected profits and losses associated with those probabilities. Although the proportion of reversals varied with the characteristics of the pairs, as found in our (1971) study, strong reversal effects were generally observed. Group judgments and decisions were even more prone to reversals than those of individuals. Because group decisions involve discussion of strategies, this result is congruent with the effects of discussion found by Hamm. Mowen and Gentry related the anchoring process thought to determine pricing responses to an anecdote provided by R. A. Kerr (1979) regarding the search for oil in the Baltimore Canyon. Kerr noted that oil companies paid $1.1 billion for the privilege of drilling despite negative reports from oil industry geochemists. He con-

[2] See our article (1973).

cluded that "Company managers apparently bid more on the basis of how large the possible trapping structures were rather than on the basis of the odds figured by the geochemists" (p. 1071).

In sum, many of the concerns raised and examined by Grether and Plott have also been investigated in other studies of preference reversals. Our purpose in reviewing these studies is not to deny the importance of the studies by Grether and Plott, Pommerehne et al., and Reilly, but rather to inform those interested in this topic about the larger body of results. In our opinion, the most striking result of these studies is the persistence of preference reversals in the face of determined efforts to minimize or eliminate them.

II. A Broader View of Preference Reversals

The inconsistency between prices and choices for risky prospects represents but one of a broad set of failings that have been attributed to the theory of rational choice. James March (1978, 1982) has identified five general problems with the theory, one of which is particularly relevant to the present discussion. According to March, the theory presumes two improbably precise guesses about the future. One is a guess about the future consequences of current actions. The other is a guess about future sentiments (i.e., preferences) with respect to those consequences.

March (1978) argued that, partly as a result of behavioral research on human information-processing limitations, the way that the rational theory deals with the first guess has been modified to incorporate principles of what Herbert Simon (1957) termed "bounded rationality." Thus economic theories now place considerable emphasis on notions of search, attention, and information costs. Aspiration levels, incrementalism, and satisficing have been described as sensible in many settings.

In contrast, March observed that although the second guess, about uncertain preferences, has so far had little effect in modifying normative theories, it poses potentially greater difficulties for these theories and their applications. He argued that limited cognitive capacity affects information processing about preferences just as it affects information processing about consequences: "Human beings have unstable, inconsistent, incompletely evoked, and imprecise goals at least in part because human abilities limit preference orderliness" (1978, p. 598).

March draws upon a rich and diverse array of observations to argue that, contrary to normative theory, preferences are neither absolute, stable, consistent, precise or exogenous (unaffected by the choices they control). The case against consistency brings us back to the topic of preference reversals. Inconsistencies between prices and choices were created on the basis of knowledge about different rules for processing the component aspects or dimensions of gambles. Since 1968, when information processing ideas began to be applied to risky choice, we have learned more about how perception and cognition determine preferences. As we have better understood those processes, it has become relatively easy, indeed almost commonplace, to produce new kinds of preference reversals. In many instances, production of reversals has been used to validate hypotheses about information processing in risky choice.

An early demonstration of the link between information processing and reversals was a study by Amos Tversky (1969). He hypothesized that, where the structure of the choice set permitted, it would be simpler and more natural to compare alternatives dimension by dimension than to evaluate the combined worth of each alternative separately (across dimensions) and then compare these overall evaluations. Tversky further hypothesized that small differences (for example, below some threshold of discrimination) would be ignored, even for an important dimension. Tversky tested and confirmed these hypotheses by creating sets of gambles in which this sort of information processing led to systematic, predictable intransitivities. Tversky's gambles contained only two dimensions, probability of winning and amount to win. For his subjects, probability was the dominant dimension, but if the difference between gambles was small, then amount to win controlled the decision. Thus, given the set of gambles a, b, c, d, and e with probabilities of 7/24, 8/24, 9/24, 10/24, and 11/24 to win $5.00, $4.75, $4.50, $4.25, and $4.00, respectively, a tended to be chosen over b, b over c, c over d, and d over e, presumably because the difference in payoffs outweighed the slight difference in probabilities within each of these pairs. However, e was typically chosen over a because of the relatively large difference in probabilities. This general finding has subsequently been replicated and extended by Rob Ranyard (1976) and by Lindman and James Lyons (1978).

The intransitivities observed by Tversky arose from the tendency of subjects to compare gambles on each dimension. If they had made holistic evaluations separately for each gamble and compared these to determine their choices, then the intransitivities would not have occurred. Comparison within dimensions is a natural way to choose among multidimensional objects. However, information is sometimes not available for each dimension, a situation that can lead to reversals. Consider, for example, the task of predicting which of two college students, A or B, would get the higher grade point average. Two test scores are available for each student, to serve as the basis for prediction. One score, English Skill, is available for both students. The other information is unique —Quantitative Ability for Student A and Need for Achievement for Student B as shown below (the means and standard deviations of each test are different but are known to the evaluator).

	Student A	Student B
Need for Achievement	–	30
English Skills	90	131
Quantitative Ability	602	–

Slovic and Douglas MacPhillamy (1974) hypothesized that commonality would cause a dimension to be weighted more heavily in determining a choice, because common information is easier to use. This, in fact, occurred and led to systematic reversals on the above problem: 75 percent of the subjects rating the students individually gave a higher grade point average to Student A. However, when these same subjects were asked to make a comparative judgment, they selected Student B 60 percent of the time (40 percent of the subjects exhibited reversals). Reversals also occurred, though less frequently, when the means and standard deviations were the same for each test.

A variety of different reversals, providing strong evidence against traditional theories of preference, have come from the work of Daniel Kahneman and Tversky (1979; Tversky and Kahneman, 1981). From their systematic observations of choices among risky alternatives, Kahneman and Tversky have deduced a number of general principles, some of which violate expected utility theory, others of which are incompatible with all existing theories of choice or preference. Kahneman and Tversky distinguished between two phases in the choice process, an early phase of editing and a subsequent phase of evaluation. The editing phase, which they have also referred to as *framing*, consists of a preliminary analysis of the available options, their possible outcomes, and the contingencies or conditional probabilities relating outcomes to acts. One function of the framing process is to organize and reformulate the alternatives so as to simplify the second phase of evaluation and choice. Much as changes in vantage point induce alternative perspectives on a visual scene, the same decision problem can be subject to many alternative frames. Whichever frame is adopted is determined in part by the external formulation of the problem and in part by the standards, habits, and personal predilections of the decision maker.

A key element of framing is the coding of outcomes. Kahneman and Tversky show that, contrary to utility theory, outcomes are typically coded as gains and losses, rather than as final states of wealth. These gains and losses are defined relative to some neutral reference point, typically, but not always, the current asset position of the decision maker. These changes are evaluated according to a value function, $v(x)$, which attaches a subjective worth to each possible outcome of a gamble, and a nonlinear probability weighting function, $\pi(p)$, which expresses the subjective importance attached to the probability of obtaining a particular outcome. The attractiveness of a gamble that offers a chance of p to obtain outcome x and a chance of q to obtain outcome y would be equal to $\pi(p)v(x)+\pi(q)v(y)$. In addition to being defined on gains and losses relative to some psychologically meaningful (neutral) reference point, the value function is steeper for losses than for gains, meaning that a given change in one's status hurts more as a loss than it pleases as a gain. Another important feature is that the function is concave above that reference point and convex below it, meaning, for example, that the subjective difference between gaining (or losing) \$10 and \$20 is greater than the difference between gaining (or losing) \$110 and \$120. Perhaps the most notable feature of the probability weighting function is the great importance attached to outcomes that will be received with certainty. Thus, for example, the prospect of losing \$50 with probability of 1.0 is more than twice as aversive as the prospect of losing the same amount with probability .5.

The way a problem is framed determines both the reference point (the zero point) of the value function and the probabilities that are evaluated. If π and v were linear functions, preferences among options would be independent of the framing of acts, outcomes, or contingencies. Because of the characteristic nonlinearities of π and v, however, normatively inconsequential changes in the frames significantly affect preferences. This is illustrated by the following pair of problems, given to separate groups of respondents.

Problem 1. Imagine that the United States is preparing for the outbreak of an unusual disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the consequences of the programs are as follows: If Program *A* is adopted, 200 people will be saved. If Program *B* is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved. Which of the two programs would you favor?

Problem 2. (Same cover story as Problem 1.) If Program *C* is adopted, 400 people will die. If Program *D* is adopted, there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die. Which of the two programs would you favor?

Although the two problems are formally identical, the preferences tend to be quite different. In a study of college students, 72 percent of the respondents chose Program *A* over Program *B* and 78 percent chose Program *D* over Program *C*. This difference can be traced to the different frames implied by the two problems. The "save lives" wording of the first problem implies that the value function's reference point is the loss of 600 lives, while the "people will die" wording of problem 2 suggests that the reference point is at no lives lost. Thus problem 1 falls in the concave gain region of the value function while problem 2 is in the convex loss region. Another study, surveying physicians and patients regarding choice of radiation vs. surgical treatments for lung cancer, produced different decisions when relevant statistics were changed from probabilities of surviving for various lengths of time after treatment to probabilities of not surviving (Barbara McNeil et al., 1982).

Another example of framing effects has been presented by Kahneman and Tversky (1982).

Problem 1. Imagine that, in addition to whatever else you have, you have been given \$200. You are now asked to choose between (A) a sure gain of \$50 and (B) a 25 percent chance of winning \$200 and a 75 percent chance of winning nothing.

Problem 2. Imagine that, in addition to whatever you have, you have been given a cash gift of \$400. You are now asked to choose between (C) a sure loss of \$150, and (D) a 75 percent chance of losing \$200 and a 25 percent chance of losing nothing.

Most people choose A over B and D over C. Yet, the options presented in the two problems are identical. There is no valid reason to prefer the gamble in one version and the sure outcome in the other. Choosing the sure gain in the first problem yields a total gain of \$200 plus \$50, or \$250. Choosing the sure loss in the second version yields the same result through the deduction of \$150 from \$400. The choice of the gamble in either problem yields a 75 percent chance of winning \$200 and a 25 percent chance of winning \$400. If the respondents to these problems took a comprehensive view of the consequences, as is assumed by theories of rational decision, they would combine the bonus with the available options and evaluate the composite outcome. Instead they ignore the bonus and evaluate the first problem as a choice between gains and the second as a choice between losses. The reversal of preferences is induced by reframing the problem.

We have used the framing and reference point notions to explain the finding that the certain loss of a stated amount of money (for example, \$50) was much more attractive when described as an insurance premium (to safeguard against a .25 chance of losing \$200) than when it was described as an alternative to playing that same gamble (see our article with Baruch Fischhoff, 1982a; see also Paul Schoemaker and Howard Kunreuther, 1979, and John Hershey and Schoemaker, 1980, for similar results).

III. Where Next?

We have presented a sample of the sorts of preference reversals that have formed our understanding of choice processes or have been created from that understanding. Those who are concerned about the possible economic implications of these phenomena have several paths to consider. One is to continue to subject these studies to the sorts of scrutiny that Grether and Plott and others have applied to the inconsistency between prices and choices. Despite the claims by Tversky and Kahneman (1981) that the effects they described are large and systematic, associated with losses of human life as well as monetary outcomes, not restricted to hypothetical questions, nor eliminated by monetary incentives, this line of research is young and there is certainly a need to test the limits and robustness of its findings.

A second path is to modify utility theory in order to accommodate as many of the behavioral anomalies as possible without abandoning the theory altogether. This has been a popular direction in recent years. A number of theorists have proposed weakening or eliminating the substitution axiom in order to accommodate the Allais paradox (Maurice Allais, 1953) and certain other violations of the traditional model (see S. H. Chew and Kenneth MacCrimmon, 1979; Peter Fishburn, 1981; Robert Weber, 1982; Hector Munera and Richard de Neufville, 1982; and Mark Machina, 1982). However, none of these revamped models can explain the framing effects described by Tversky and Kahneman (1981) or the preference reversals among *P* bets and \$ bets. Indeed, Machina acknowledged that, "to the extent that preference reversals are found to be systematic and pervasive, the behavioral model presented here must either be generalized or replaced" (p. 308).

A third path to follow, and one that we would advocate, is to accept the reality of preference reversals and related information-processing phenomena, and to explore their implications for important social and economic behaviors. We have begun to do this with regard to problems of societal risk management and programs for informing the public about risk (see our study, with Fischhoff, 1982b). Similarly, March (1978, 1982), whose critique went far beyond information processing to encompass complex strategic and social motivations, has urged that a conception of preference that respects the "intelligence of ambiguity" be incorporated into what he calls "the engineering of choice." He identified a number of conceptual problems that need to be addressed by choice theorists and optimization problems that need to be considered by choice engineers.

In a narrower but nonetheless important vein, Hershey, Kunreuther, and Schoemaker (1982) have demonstrated biases in utility functions caused by information processing effects. They showed that methods for assessing utilities, varying in normatively inconsequential ways, produced very different utility functions, posing both practical and theoretical problems for those concerned with assessing people's risk preferences. Donald Wehrung, MacCrimmon, and K. Brothers (1980) obtained similar inconsistencies with business executives, leading them to question the use of utility theory as a management tool. A more general analysis of the difficulties of assessing preferences has been presented by Fischhoff and ourselves (1980).

Fischhoff et al. argue that the strong effects of framing and information-processing considerations make elicitation methods major forces in shaping the expression of one's personal values.

Robin Gregory (1982) investigated a number of different approaches for estimating the value of nonmarket goods such as air and water quality, protection of threatened environments and species, and access to uninhabited views. He examined two measures of economic value; one based on an individual's willingness to pay to obtain or retain a good and the other based on the amount of compensation demanded if it is relinquished. He found that both methods were subject to sizable framing and information-processing effects.

Richard Thaler (1980) has drawn upon the reference point and framing notions of Kahneman and Tversky to explain a number of "economic illusions" that cause consumer behavior to deviate from the predictions of normative models. Included in his analysis were the overweighting of out-of-pocket costs relative to opportunity costs (foregone gains), the failure to ignore sunk costs, and the effects of psychic regret on such diverse areas as health care delivery decisions and vacation planning. Thomas Russell and Thaler (1982) argued that departures from rationality due to information-processing effects are unlikely to disappear in competitive markets. Kenneth Arrow (1982) underscored this argument by pointing out a number of failures of the rational model in insurance, securities, and futures markets that he feels are directly interpretable in terms of effects such as those linked to preference reversals and framing.

IV. Conclusion

This review has attempted to show how preference reversals fit into a larger picture of information-processing effects that, as a whole, pose a collective challenge to preference theories far exceeding that from reversals alone. These effects seem unlikely to disappear, even under rigorous scrutiny. Moreover, anything less than a radical modification of traditional theories is unlikely to accommodate these phenomena. We urge economists not to resist these developments but, instead, to examine them for insights into the ways that decisions are made and the ways that the practice of decision making can be improved.

REFERENCES

Allais, Maurice, "Le Comportement de l'Homme Rationnel devant le Risque: Critiques des Postulats et Axiomes de l'Ecole Americaine," *Econometrica*, October 1953, *21*, 503–46.

Arrow, Kenneth J., "Risk Perception in Psychology and Economics," *Economic Inquiry*, January 1982, *20*, 1–9.

Chew, S. H. and MacCrimmon, Kenneth, "Alpha Utility Theory, Lottery, Composition, and the Allais Paradox," Working Paper No. 686, University of British Columbia, 1979.

Fischhoff, Baruch, Slovic, Paul and Lichtenstein, Sarah, "Knowing What You Want: Measuring Labile Values," in Tom Wallsten, ed., *Cognitive Processes in Choice and Decision Behavior*, Hillsdale: Erlbaum, 1980.

Fishburn, Peter C., "Transitive Measurable Utility," Economic Discussion Paper No. 224, Bell Laboratories, Murray Hill, 1981.

Gregory, Robin S., "Valuing Non-Market Goods: An Analysis of Alternative Approaches," unpublished doctoral dissertation, University of British Columbia, June 1982.

Grether, David M. and Plott, Charles R., "Economic Theory of Choice and the Preference Reversal Phenomenon," *American Economic Review*, September 1979, *69*, 623–38.

Hamm, Robert MacGowan, "The Conditions of Occurrence of the Preference Reversal Phenomenon," unpublished doctoral dissertation, Harvard University, 1979.

Hershey, John C., Kunreuther, Howard and Schoemaker, Paul J. H., "Bias in Assessment Procedures for Utility Functions," *Management Science*, August 1982, *28*, 936–54.

______ **and Schoemaker, Paul J. H.,** "Risk Taking and Problem Context in the Domain of Losses: An Expected Utility Analysis," *Journal of Risk and Insurance*, March 1980, *47*, 111–32.

Kahneman, Daniel and Tversky, Amos, "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, March 1979, *47*, 263–91.

______ **and** ______, "The Psychology of Preferences," *Scientific American*, January 1982, *246*, 160–73.

Kerr, R. A., "Petroleum Exploration: Discouragement about the Atlantic Outer Continental Shelf Deepens," *Science*, June 8, 1979, *204*, 1069–72.

Lichtenstein, Sarah and Slovic, Paul, "Reversals of Preference Between Bids and Choices in Gambling Decisions," *Journal of Experimental Psychology*, January 1971, *89*, 46–55.

______ **and** ______, "Response-Induced Reversals of Preference in Gambling: An Extended Replication in Las Vegas," *Journal of Experimental Psychology*, November 1973, *101*, 16–20.

Lindman, Harold R., "Simultaneous Measurement of Utilities and Subjective Probabilities," unpublished doctoral dissertation, University of Michigan, 1965.

______, "Inconsistent Preferences Among Gambles," *Journal of Experimental Psychology*, May 1971, *89*, 390–97.

______ **and Lyons, James,** "Stimulus Complexity and Choice Inconsistency Among Gambles," *Organizational Behavior and Human Performance*, April 1978, *21*, 146–59.

McNeil, Barbara J. et al., "On the Elicitation of Preferences for Alternative Therapies," *New England Journal of Medicine*, May 1982, *306*, 1259–62.

Machina, Mark, J., "'Expected Utility' Analysis Without the Independence Axiom," *Econometrica*, March 1982, *50*, 277–323.

March, James G., "Bounded Rationality, Ambiguity, and the Engineering of Choice," *Bell Journal of Economics*, Autumn 1978, *9*, 587–608.

______, "Theories of Choice and the Making of Decisions," presented at the Annual Meeting of the American Association for the Advancement of Science, Washington, D.C., January 5, 1982.

Mowen, John C. and Gentry, James W., "Investigation of the Preference-Reversal Phenomenon in a New Product Introduction Task," *Journal of Applied Psychology*, No. 6, 1980, *65*, 715–22.

Munera, Hector A. and de Neufville, Richard, "A Decision Analysis Model when the Substitution Principle is not Acceptable," presented to the First International Conference on Foundations of Utility and Risk Theory, Oslo, July 1982.

Pommerehne, Werner W., Schneider, Friedrich and Zweifel, Peter, "Economic Theory of Choice and the Preference Reversal Phenomenon: A Reexamination," *American Economic Review*, June 1982, *72*, 569–74.

Ranyard, Rob H., "Elimination by Aspects as a Decision Rule for Risky Choice," *Acta Psychologica*, August 1976, *40*, 299–310.

Reilly, Robert J., "Preference Reversal: Further Evidence and Some Suggested Modifications in Experimental Design," *American Economic Review*, June 1982, *72*, 576–84.

Russell, Thomas and Thaler, Richard, "The Relevance of Quasi-Rationality in Competitive Markets." Working Paper, Graduate School of Business and Public Administration, Cornell University, 1982.

Schoemaker, Paul J. H. and Kunreuther, Howard C., "An Experimental Study of Insurance Decisions," *Journal of Risk and Insurance*, December 1979, *46*, 603–18.

Simon, Herbert A., *Models of Man*, New York: Wiley, 1957.

Slovic, Paul and Lichtenstein, Sarah, "Relative Importance of Probabilities and Payoffs in Risk Taking," *Journal of Experimental Psychology Monograph*, November 1968, Pt. 2, *78*, 1–18.

______, **Fischhoff, Baruch and Lichtenstein, Sarah,** (1982a) "Response Mode, Framing, and Information-Processing Effects in Risk Assessment," in Robin Hogarth, ed., *New Directions for Methodology of Social and Behavioral Science: Question Framing and Response Consistency*, San Francisco: Jossey-Bass, 1982.

______, ______, **and** ______, (1982b) *Regulation of Risk: A Psychological Perspective*, Decision Research Report 82-11, Decision Research: Eugene, 1982.

______ and **MacPhillamy, Douglas,** "Dimensional Commensurability and Cue Utilization in Comparative Judgment," *Organizational Behavior and Human Performance*, April 1974, *11*, 172–94.

Thaler, Richard, "Toward a Positive Theory of Consumer Choice," *Journal of Economic Behavior and Organization*, March 1980, *1*, 39–60.

Tversky, Amos, "Intransitivity of Preferences," *Psychological Review*, January 1969, *76*, 31–48.

______ and **Kahneman, Daniel,** "The Framing of Decisions and the Psychology of Choice," *Science*, January 1981, *211*, 453–58.

Weber, Robert J., "The Allais Paradox, Dutch Auctions, and Alpha-Utility Theory," Working Paper, J.L. Kellogg Graduate School of Management, Northwestern University, September 1982.

Wehrung, Donald, A., MacCrimmon, Kenneth, R. and Brothers, K. M., "Utility Measures: Comparisons of Domains, Stability, and Equivalence Procedures," Working Paper No. 603, University of British Columbia, February 1980.

Reprinted from

THE AMERICAN ECONOMIC REVIEW

END

FILMED

1-85

DTIC